ARMY RESEARCH LABORATORY

Proceedings of the First Annual U.S. Army Conference on Applied Statistics, 18–20 October 1995

Barry Bodt
Proceedings Chairman

Hosted by:
U.S. ARMY RESEARCH LABORATORY

Cosponsored by:
U.S. ARMY RESEARCH LABORATORY
U.S. MILITARY ACADEMY
U.S. ARMY RESEARCH OFFICE
WALTER REED ARMY INSTITUTE OF RESEARCH
NATIONAL INSTITUTE OF STANDARDS AND TECHNOLOGY
TRADOC ANALYSIS CENTER-WSMR

ARL-SR-43 August 1996

NOTICES

Destroy this report when it is no longer needed. DO NOT return it to the originator.

Additional copies of this report may be obtained from the National Technical Information Service, U.S. Department of Commerce, 5285 Port Royal Road, Springfield, VA 22161.

The findings of this report are not to be construed as an official Department of the Army position, unless so designated by other authorized documents.

The use of trade names or manufacturers' names in this report does not constitute indorsement of any commercial product.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY *(Leave blank)*	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	August 1996	Final, 18-20 October 1995

4. TITLE AND SUBTITLE
Proceedings of the First Annual U.S. Army Conference on Applied Statistics, 18-20 October 1995

5. FUNDING NUMBERS
PR: 1L162618AH80

6. AUTHOR(S)
Barry Bodt

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
U.S. Army Research Laboratory
ATTN: AMSRL-SC-S
Aberdeen Proving Ground, MD 21005-5067

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAMES(S) AND ADDRESS(ES)

10.SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution is unlimited.

12b. DISTRIBUTION CODE

13. ABSTRACT ***(Maximum 200 words)***

The first U.S. Army Conference on Applied Statistics was held 18-20 October 1995 at the headquarters of the U.S. Army Research Laboratory in Adelphi, MD. The conference was cosponsored by the U.S. Army Research Laboratory, the U.S. Military Academy, the Army Research Office, the Walter Reed Army Institute of Research, the National Institute of Standards and Technology, and the TRADOC Analysis Center-WSMR. Papers given at the conference addressed the development of new statistical techniques, application of existing methodologies to Army problems, and panel discussion of statistical challenges in an Army setting. This report is a compilation of available papers offered at the conference.

14. SUBJECT TERMS	15. NUMBER OF PAGES
applied statistics, experimental design, statistical inference	157
	16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	UL

INTENTIONALLY LEFT BLANK.

FOREWORD

The first U.S. Army Conference on Applied Statistics was held 18–20 October 1995, at the headquarters of the U.S. Army Research Laboratory (ARL) in Adelphi, MD. The conference was cosponsored by ARL, the U.S. Army Research Office (ARO), the U.S. Military Academy (USMA), the Training and Doctrine Command (TRADOC) Analysis Center-White Sands Missile Range (TRAC-WSMR), the Walter Reed Army Institute of Research (WRAIR), and the National Institute of Standards and Technology (NIST). The U.S. Army Conference on Applied Statistics is successor to the U.S. Army Conference on the Design of Experiments, a historic series of meetings that formally concluded in 1994 after 40 years of service to the Army. Today's Army faces challenges that are far ranging and encompass many topics in which probability and statistics have a contribution to make, in addition to experimental design. This new conference reflects a broadening of scope with the goal to promote the practice of statistics in the solution of diverse Army problems.

This first conference offered much with respect to that goal. Toward statistical education, the conference was preceded with a short course, "Tree-Structured Methods," given by Professor Wei-Yin Loh of the University of Wisconsin at Madison. Several distinguished speakers, from government, industry, and academia, spoke during invited general sessions: Professor William J. Conover, Texas Tech University; Dr. John W. Green, DuPont; Mr. Roy Reynolds, Director, TRADOC Analysis Center-WSMR; Professor Max Woods, Naval Postgraduate School; and Professor James E. Gentle, George Mason University. Contributed talks developed new methodology, detailed successful applications, or requested guidance from a panel of experts in attacking an Army problem that had resisted standard statistical approaches. A special session was devoted to the unique difficulties of advanced warfighting experiments.

The Executive Board for the conference recognizes several individuals for their contributions to conference details: Dr. Douglas Tang, WRAIR; Dr. Mark Vangel, NIST; Dr. Eugene Dutoit, U.S. Army Infantry School (AIS); and Dr. Paul Deason, TRAC-WSMR. Drs. Barry Bodt and Malcolm Taylor, ARL, are recognized for organizing and hosting the meeting. Dr. Bodt oversaw the publishing of the Proceedings. Special thanks is due Mrs. Patricia Cizmadia of the Protocol Office, ARL, who, with Mrs. Tammy Bassford and Ms. Karen Moore assisting, served as site coordinator for the conference.

INTENTIONALLY LEFT BLANK.

TABLE OF CONTENTS*

* This Table of Contents contains only the papers that appear in the Proceedings.

Conference Agenda

Wednesday, 18 October 1995

0900 - 0930 CALL TO ORDER

OPENING REMARKS: John Lyons, Director, ARL

0930 - 1200 GENERAL SESSION I

Chair: Malcolm S. Taylor, ARL

0930 - 1030 KEYNOTE ADDRESS

STATISTICS IN THE YEAR 2000-WHERE ARE WE GOING?
W. J. Conover, Texas Tech University

1030 - 1100 Break

1100 - 1200 HYPOTHESIS TESTING VS REGRESSION MODELING IN TOXICOLOGY
John W. Green, DuPont

1200 - 1330 Lunch

1330 - 1500 CONTRIBUTED SESSION I (in parallel with Clinical Session I)

Chair: David Webb, ARL

INFERENCE ON A COMMON MEAN IN INTERLABORATORY STUDIES
Mark G. Vangel, ARL and NIST

NONLINEAR MIXED EFFECTS METHODOLOGY FOR RHYTHMIC DATA
R. John Weaver* and M. N. Brunden, Upjohn Laboratories

PRIORITY SELECTION TO MINIMIZE EXPECTED LOSS
Ann E. M. Brodeen, ARL
Douglas H. Frank,* Indiana University of Pennsylvania

1330 - 1500 CLINICAL SESSION I (in parallel with Contributed Session I)

Chair: James E. Gentle, George Mason University

Panel: W. J. Conover, Texas Tech University
John W. Green, DuPont
James R. Thompson, Rice University

TWO STAGE EXPERIMENT TO DETERMINE RESPONSE (INCREASE IN HUMAN PERFORMANCE) DUE TO A CONTROLLED EXPERIMENT
Lou Delattre, U.S. Army Materiel Systems Analysis Agency

STATISTICAL ANALYSIS OF LETHALITY EXCHANGE RATIO
David Grant and Albert W. Van Horn,* U.S. Army Tank-automotive and Armaments Command

CHARACTERIZING NONMONOTONIC RESPONSE CURVES: BALLISTIC LIMITS IN THE PRESENCE OF SHATTER GAP
Albert Chang* and Barry A. Bodt,* ARL

1500 - 1515 Break

1515 - 1615 CONTRIBUTED SESSION II (in parallel with Contributed Session III)

Chair: Marc Elliott, RAND

METHODOLOGY FOR THE CURVE FITTING OF NONLINEAR RIDE CURVES
Andrew W. Harrell, U.S. Army Engineer Waterways Experiment Station

AN UNCONVENTIONAL APPROACH TO CURVE FITTING FROM STATISTICAL DATA
Mel Brown, ARO

1515 - 1615 CONTRIBUTED SESSION III (in parallel with Contributed Session II)

Chair: Robert Burge, WRAIR

INDIVIDUAL BIOEQUIVALENCE: THE BIOEQUIVALENCE SENSITIVITY RATIO COMPARES CRITERION OF BIOAVAILABILITY MERIT
Marshall N. Brunden* and Thomas J. Vidmar, Upjohn Laboratories

SOME STATISTICAL ISSUES FOR IMPROVED RELIABILITY AND MAINTAINABILITY OF AGING STRUCTURES
Asit P. Basu, University of Missouri-Columbia

Thursday, 19 October 1995

0830 - 0930 GENERAL SESSION II

Chair: Bernard Harris, University of Wisconsin-Madison

HOW KNOWLEDGE ABOUT WEAR-OUT CAN BE USED TO REDUCE TIME-TRUNCATED DEMONSTRATION TESTING
W. Max Woods, Naval Postgraduate School

0930 - 1000 Break

1000 - 1130 CONTRIBUTED SESSION IV

Chair: Robert Launer, ARO

APPLICATIONS OF SUBJECTIVE PROBABILITY TO TACTICAL DECISION MAKING: A NEW WEAPON IN THE INFORMATION WAR
MAJ David H. Olwell, USMA

A BAYESIAN PARETO ANALYSIS FOR SYSTEM OPTIMIZATION OF NASA's SPACE STATION
James R. Thompson* and Roxy D. Walsh, Rice University

DISCERNING A FAIR FIGHT ON DISSIMILAR SIMULATORS
Lorrie L. Hoffman, University of Central Florida

1130 - 1300 Lunch

1300 - 1400 GENERAL SESSION III

Chair: Paul Deason, TRAC-WSMR

THE ROLE OF ANALYSIS IN DEVELOPING THE FUTURE FORCE
Roy Reynolds, Director, TRADOC Analysis Center-WSMR

1400 - 1415 Break

1415 - 1615 SPECIAL SESSION ON ADVANCED WARFIGHTING EXPERIMENTS (AWE) (in parallel with Contributed Session V)

Chair: Eugene Dutoit, U.S. AIS

INTRODUCTION TO THE SPECIAL SESSION ON AWEs AND HOW IT RELATES TO THIS CONFERENCE
Eugene Dutoit, U.S. AIS

EVOLUTION OF THE MODEL-TEST-MODEL CONCEPT FOR USE IN OPERATIONAL TESTING AND ADVANCED WARFIGHTING EXPERIMENTS
Bryson McCool, Jerry Lyman, and LTC John Ferguson, TRAC-WSMR

PERFORMANCE BASED METRICS FOR THE DIGITIZED BATTLEFIELD
Jock Grynovicki,* Dennis Leedom, Michael Golden, Josephine Wojciechowski, ARL; Gene Dutoit, U.S. AIS

1415 - 1615 CONTRIBUTED SESSION V (in parallel with Special Session on AWE)

Chair: David Cruess, Uniform Services University of the Health Sciences

ON THE PERFORMANCE OF WEIBULL LIFE TESTING PROCEDURES BASED ON EXPONENTIAL LIFE TESTING DESIGNS
Francisco J. Samaniego, University of California-Davis

VALIDATING JUDGEMENTS AND DECISIONS FOR INPUT TO COMBAT SIMULATIONS: AN STF EXAMPLE THAT QUANTIFIES THE VALUE OF SCOUT RECONNAISSANCE
Clairice Veit, RAND

PENETRATION AND DEFLATION ALGORITHMS FOR TIRE VULNERABILITY
Linda L. C. Moss,* Ricky L. Grote, and Edwin O. Davisson, Jr., ARL

USING REAL-WORLD AND SIMULATION DATA TO ESTIMATE A LOCATION PARAMETER
Dennis E. Smith, Desmatics, Inc.

Friday, 20 October 1995

0830 - 0930 GENERAL SESSION IV

Chair: Edward Wegman, George Mason University

MULTICRITERIA OPTIMIZATION IN STATISTICAL DESIGN AND ESTIMATION
James E. Gentle, George Mason University

0930 - 0945 Break

0945 - 1115 CONTRIBUTED SESSION VI

Chair: Carl Russell, TEC

THE EFFECTS OF A COMPUTER-AIDED TELEOPERATION TECHNOLOGY ON OPERATOR WORKLOAD AND PERFORMANCE OF CONCURRENT TASKS
Monica M. Glumm and Jock O. Grynovicki,* ARL

PROCESS-ORIENTED BASIS REPRESENTATIONS FOR MULTIVARIATE SPC
Russell Barton and David R. Gonzalez-Barreto, Pennsylvania State University

CONTRIBUTED SESSION VI *cont.*

A MARKOVIAN MODEL FOR GROWTH AND BRANCHING OF STABLE CRACKS IN METALS
Luis A. de Bejar, U.S. Army Engineer Waterways Experiment Station

1115 - 1130 Break

1130 - 1230 CLOSING REMARKS

Chair: Barry Bodt, ARL

1230 ADJOURN

NONLINEAR MIXED EFFECTS METHODOLOGY FOR RHYTHMIC DATA

R. J. Weaver and M.N. Brunden
Pharmacia & Upjohn, Inc., Kalamazoo, Michigan 49001

ABSTRACT

We develop methodology for a mixed effects Cosinor model suitable for analyzing rhythmic data. None of the currently used procedures for nonlinear mixed effects models can be directly applied to the Cosinor model. Our approach combines ideas from the areas of time series analysis and mixed effects model methodology, and addresses the inherent limitations of the current procedures, as well as new problems encountered when combining the methodologies.

INTRODUCTION

In many biological investigations data on an endpoint of interest is collected repeatedly over time for each of several individuals, which in turn may be part of a between individual experimental design. Biological time series of this type typically exhibit rhythmic behavior. As is common with biological data, there may be significant variation among individuals in these rhythm characteristics.

This experimental setup suggests using a random or mixed effects model, where a common functional form is assumed for each individual, but some or all of the parameters are considered to vary among the individuals. It is then of interest to estimate the group parameters (fixed effects) and their covariance matrix, and perhaps make comparisons between them. When the period of the rhythm is unknown, most commonly used models are nonlinear in their parameters.

Various methods are proposed in the literature for nonlinear mixed effects model, but there are several unique aspects to this particular problem that preclude using these methods directly. Ordinary nonlinear least squares estimation is difficult to use for these models due to multiple local minima. As an alternative, periodogram based estimators can be used.

Again due to the nature of the model, nonlinear mixed effects methodology using the usual Taylor's series approximation to the expected response do not work well. Problems also occur when the data are pooled together to estimate the fixed effects, as in Vonesh and Carter's EGLS methodology. These problems will be examined in detail, and a new two-stage methodology is proposed. This methodology is shown to perform well not only in rhythmic data models, but also in general nonlinear mixed effects models from pharmacokinetics and growth curve problems.

THE MIXED EFFECTS COSINOR MODEL

THE GENERAL MIXED EFFECTS MODEL

We will consider a general form of the mixed effects model, similar to that described by Lindstrom and Bates[1] .

Assumption A1. A similar functional form is assumed for each of the N individuals, that is

where $$y_i = f(\phi_i, X_i) + e_i \qquad i = 1, 2, \ldots, N$$

y_i is a $n_i \times 1$ vector of observations,
ϕ_i is a $p \times 1$ vector of (unknown) parameters for individual i,
X_i is the $n_i \times p$ within individual design matrix,

$f(\phi_i, X_i)$ is the $n_i \times 1$ expected value of the response at ϕ_i, X_i ,
ε_i is random error, assumed to be $N(0, \Sigma_i)$.

Assumption A2. Each individual's true parameter vector can be expressed as

with $$\phi_i = A_i \alpha + B_i b_i$$

A_i a $p \times q$ between individual design matrix,
α a $q \times 1$ vector of fixed effects,
B_i a $p \times r$ design matrix indicating the random parameters,
b_i a $r \times 1$ vector of random effects, for which we will assume $b_i \sim N(0, \Psi)$

This model setup includes growth models, random coefficient regression models, population pharmacokinetic models and repeated measures models as special cases. The fixed effects α can be interpreted as the group mean parameters, and in the case of a single group are sometimes referred to as the population parameters. The b_i are interpreted as the i^{th} individual's parameters deviation from the group or population mean parameter vector. Our main interest is in estimating α and the unique elements of the covariance matrix Ψ. Depending on the experimental situation, it may also be of interest to estimate the individual parameter vectors ϕ_i and/or σ^2.

Methodologies for the case when the within individual model is linear in its parameters have been developed by, among othere, Laird and Ware [2], Jennrich and Schluchter [3], and Vonesh and Carter [4]. For the nonlinear case, Steimer [5], Racine-Poon [6], Sheiner and Beal [7], Lindstrom and Bates [8], and Vonesh and Carter [8] are useful references.

THE WITHIN INDIVIDUAL MODEL - THE COSINOR MODEL

For the within individual portion of the analyses, we will use a model proposed for the analysis of biological rhythms by Halberg, Tong and Johnson [9], called the *Cosinor* model. Consider a time series y_t, $t = 1, 2, \ldots, n$, where

$$y_t = \alpha_0 \cos(\omega_0 t) + \beta_0 \sin(\omega_0 t) + \epsilon_t = A_0 \cos(\omega_0 t + \theta_0) + \epsilon_t.$$

where the errors ϵ_t are assumed to be independent with E (ϵ) = 0 and Var (ϵ) = σ^2 for all t. In the second parameterization, A_0 is the amplitude, ω_0 is the frequency and θ_0 is the phase of the cosine curve. Further details have been given in Halberg, et. al. [10], Nelson, et. al.[11] and Bingham, et. al.[12] . This model has been extensively used and reported in the chronobiology literature, and computer programs for its implementation have been published by Monk and Fort [13] and Vokac[14] .

PROBLEMS WITH CURRENT NME METHODOLOGY

NONLINEAR LEAST SQUARES ESTIMATION FOR THE COSINOR MODEL

The Cosinor model with unknown frequency is not linear in its parameters, nor can it be made linear by a transformation of the data. The most common approach to parameter estimation in such models is nonlinear least squares. These methods involve some type of iterative search procedure, beginning at a an initial guess for the parameters and proceeding until a specified convergence criterion is met. Difficulties arise for the Cosinor model because the objective function to be minimized, the residual sum of squares

$$Q_n(\alpha, \beta, \omega) = \sum_{t=1}^{n} [y_t - \alpha \cos(\omega t) - \beta \sin(\omega t)]^2$$

has many local minima, maxima and inflection points. This problem has been discussed in some detail by Rice and Rosenblatt [15], who state that the local minima occur with a separation with respect to the frequency of about n . The main implication is that convergence to the global mimimum is very sensitive to the choice of starting values.

We will illustrate this difficulty using the example by Rice and Rosenblatt [15]. The model considered is the Cosinor model with $\alpha_0 = 1.0$, $\beta_0 = 0.0$, (or alternatively, $A_0 = 1.0$, $\theta_0 = 0.0$) $\omega_0 = 0.5$ and n = 100. To examine the problem quantitatively, a single realization of this model with Gaussian noise of mean 0 and variance 1 was randomly generated. Using this data set, the parameters were estimated by nonlinear least squares. The calculations were made using the IMSL subroutine RNLIN, which utilizes a modified Levenberg-Marquardt algorithm. The default convergence parameters of IMSL were used. To examine the dependence of obtaining a good least squares fit on the starting values, we fit the model to this set of data 100 times, each time with different starting values. In each replication, the starting value for A_0 was set to 1.0 and starting values for θ_0 and ω_0 were randomly generated from the Uniform distributions $(-\pi, \pi)$ and (0.3, 0.7). This corresponds to about the level precision in starting values that might be obtained by "eyeballing" the data. The objective function for this set of data and range of parameters is shown graphically in Figure 1. As expected, it quite rough and displays numerous local minima, maxima and inflection points.

With these randomly generated starting values, the procedure converged to or stopped near the global minimum only 15 times out of the 100 replications. One of the more common problems was with the algorithm becoming stuck in extremely "flat" regions of the objective function, and failing to meet the convergence criteria.The choice of starting value for ω is especially critical. When the starting value was more than about .05 away from the true value, the algorithm would always converge to a local extrema rather than the global minimum. This is not surprising based on the shape of the objective function's surface. The starting values must fall within or near the long, narrow depression centered on $\omega = 0.5$ or there is little chance of success.

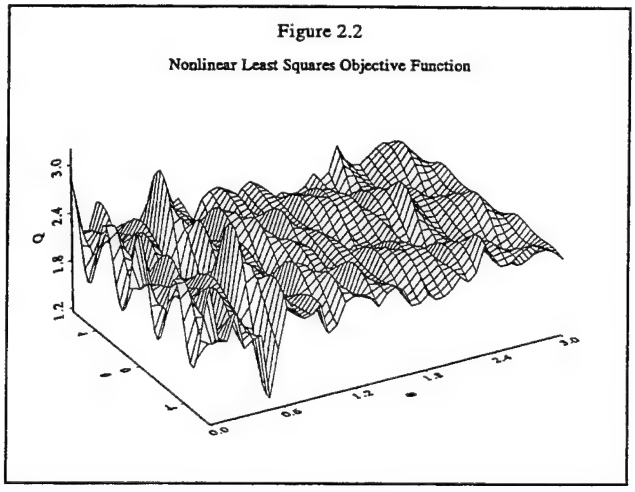


Figure 1

Another series of fits was performed with the starting values for ω generated from the Uniform distribution (.45, .55), and starting values for θ and A generated as before. This corresponds to essentially knowing the true frequency, or using a good estimate obtained from some type of independent preliminary analysis of the data. In this best case scenario, the proper estimates were obtained 92 times out of 100.

These examples indicate that it is extremely risky to rely on just a single set of starting values, even if they have been well chosen. The use of something like the GRID option of PROC NLIN of SAS, or random selection of starting values over a selected range will greatly improve the chances of finding the global minimum. Overall, these problems make the use of nonlinear least squares estimation troublesome for the Mixed Effects Cosinor model, where we have individuals with varying true parameter values.

TAYLOR'S SERIES APPROXIMATION

Both the Sheiner and Beal (NONMEM) and Lindstrom-Bates procedures make use of a Taylor's series approximation to the expectation function. This essentially utilizes a linear function to approximate a nonlinear function in a region of the true parameters. While it works well for many nonlinear functions, intuitively it does

not seem reasonable for periodic functions such as the cosine. To illustrate how it can be inadequate, we will consider the basic model with all parameters considered random, i.e. when $A_i = B_i = I_3$. The model, suppressing the subscript i, can then be written as

$$y_t = (\alpha_0 + b_1)\cos[(\omega_0 + b_3)t] + (\beta_0 + b_2)\sin[(\omega_0 + b_3)t] + e_t$$

Taking the derivative of the expectation with respect to the parameters $(\alpha_0, \beta_0, \omega_0)$ and evaluating at $b = (0, 0, 0)^t$ we get

$$y_t \approx (\alpha_0 + b_1)\cos(\omega_0 t) + (\beta_0 + b_2)\sin(\omega_0 t)$$
$$+ b_3 t(\beta_0 \cos(\omega_0 t) - \alpha_0 \sin(\omega_0 t))$$

The third term of this approximation involves the value t, and for a nonzero realization of b_3, it will increase in magnitude as t increases. This results in the approximation worsening as the length of the data series gets longer, which is a very undesirable property. This is shown graphically in Figure 2. The model shown has $\alpha_0 = \beta_0 = 25$ and $\omega_0 = 2\pi/24$, with the vector b randomly generated from a Normal distribution with mean zero and covariance

$$\Psi = \begin{pmatrix} 1 & .1 & .01 \\ .1 & 1 & .01 \\ .01 & .01 & .001 \end{pmatrix}$$

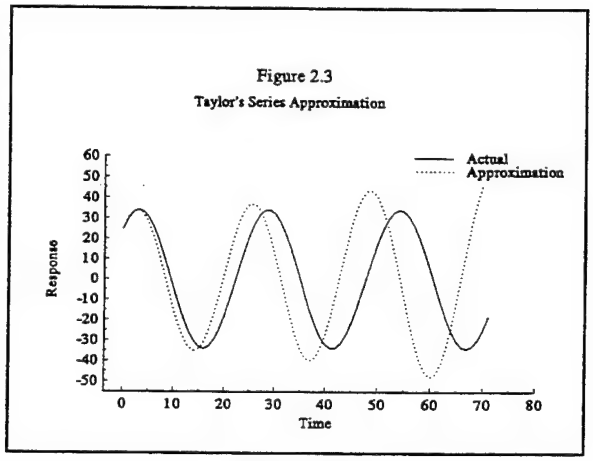


Figure 2

The actual model was then calculated for three periods, and graphed with the Taylor's series approximation superimposed on it. The approximation is quickly diverging from the true model, even for this relatively short series.

NAIVE POOLED DATA APPROACH

The Naive Pooled Data approach has been used for estimation of population parameters, and is also the first step of the Vonesh and Carter noniterative algorithm for nonlinear mixed effects models. This procedure pools the data from all individuals and estimates the population parameters.When the underlying model is the Cosinor, the Naive Pooled Data approach can result in problems if the individuals do not all share the same true phase. This well known problem of phase differences in biological time series has been discussed by Sollberger [16]. Simply put, if phases differ among individuals, degrees of cancellation will occur if the data are pooled. This phenomena is sometimes referred to as interference. The most extreme case occurs when two data series have phases differing by π radians (180 degrees), when the cancellation is total and a horizontal line would be fit to the data.

As an example of this difficulty, we generated 10 data series of 100 points each. Each series had an identical amplitude of 8.0 and frequency of 0.5, but a randomly generated phase from a Uniform distribution on the interval $(0,2\pi)$. To this signal was added randomly generated Gaussian noise with zero mean and variance of one. Each individual series had a substantial signal/noise ratio, and was clearly rhythmic to the naked eye. When the ten series were pooled and the Cosinor model was fit, this evidence of rhythmicity was masked, giving an estimated amplitude of 0.2095, not much different than a horizontal straight line.

NEW METHODOLOGY FOR THE MIXED EFFECTS COSINOR MODEL

INTRODUCTION

It is evident that no currently proposed methodology can be directly applied to the Mixed Effects Cosinor Model. We will introduce new methodology that is flexible enough to be used for the general mixed effects model, while at the same time is more appropriate when our within-individual model is the Cosinor. Our strategy is to use a two-stage procedure, where the first stage is to estimate the individual parameter vectors. In our case, we will use the Adjusted Composite Periodogram estimators of Weaver [17] for the Cosinor model . These estimators are chosen because unlike the nonlinear least squares estimators, they are essentially unbiased. For other types of problems, the first stage could be maximum likelihood, least squares, or some other type of estimation. Once we have obtained individual parameter estimates, the second stage is essentially a linear mixed effects problem, but with the additional information of estimated covariance matrices for each individual parameter vector. We will propose a 4 step noniterative procedure, which is similar to Vonesh and Carter[4] in that it utilizes estimated generalized least squares for the fixed effects and method of moments for the random effects covariance matrix Ψ.

FIRST STAGE - PARAMETER ESTIMATION

Recall that the true parameter vector for individual i is given by $\phi_i = A_i \alpha + B_i b_i$. We now add :

Assumption A3. For each individual, we can obtain an estimate of its parameter vector, denoted by $\hat{\phi}_i$, which has covariance matrix C_i. We also obtain an estimate of this covariance, which we call $\hat{C}_i$. The matrix $\hat{C}_i$ will typically be a function of $\hat{\phi}_i$. We assume that

$$\hat{\phi}_i \mid b_i \sim N (\phi_i , C_i) .$$

For many types of models, these estimates could be the maximum likelihood estimate and its asymptotic covariance matrix, or nonlinear least squares estimates.. It can be shown that the marginal distribution of the parameter estimates is

$$\hat{\phi}_i \sim N (A_i \alpha , C_i + B_i \Psi B_i^t) .$$

SECOND STAGE

Step 1. Initial estimation of the fixed effects parameters. We will first estimate the fixed effects α, assuming the random effects are equal to zero. This can be done by minimizing

$$Q_n = \sum_{i=1}^{N} (\hat{\phi}_i - A_i \alpha) \hat{C}_i^{-1} (\hat{\phi}_i - A_i \alpha)$$

which gives the usual generalized least squares estimate

$$\hat{\alpha} = \left(\sum_{i=1}^{N} A_i^t \hat{C}_i^{-1} A_i \right)^{-1} \sum_{i=1}^{N} A_i^t \hat{C}_i^{-1} \hat{\phi}_i .$$

Step 2. Estimation of the random effects. Using these estimates of α, we can compute the residuals. These

residuals can then be fit to a random coefficients regression model to estimate the b_i. The model is written as

$$\tilde{e}_i = B_i b_i + \epsilon_i \qquad \epsilon_i \sim N(0, C_i)$$

and the usual estimates are the generalized least squares estimates

$$\hat{b}_i = \left(B_i^t C_i^{-1} B_i\right)^{-1} B_i^t C_i^{-1} \tilde{e}_i .$$

The variance of the estimate is

$$Var(\hat{b}_i) = \Psi + (B_i^t C_i^{-1} B_i)^{-1}$$

To estimate the above quantities, we can replace C_i with $\hat{C}_i$.

Step 3. Estimation of the Random Effects Covariance Matrix Ψ. An estimate for Ψ will be obtained using the Method of Moments. Construct the following matrices

$$\tilde{B} = \left(\hat{b}_1\ \hat{b}_2 \ldots \hat{b}_N\right)^t \qquad \tilde{A} = \left(a_1\ a_2 \ldots a_N\right)^t$$

$$S_{bb} = \tilde{B}^t \left(I_N - \tilde{A} \left(\tilde{A}^t \tilde{A} \right)^{-1} \tilde{A}^t \right) \tilde{B}$$

where the vectors a_i are group indicators. S_{bb} is just the sample variance-covariance matrix for the b_i corrected for the between individual effects. Then, by equating S_{bb} to its expected value, we get

$$\hat{\Psi} = \left(S_{bb} - \sum_{i=1}^{N} (1 - a_i^t (\tilde{A}^t \tilde{A})^{-1} a_i)(B_i^t \hat{C}_i^{-1} B_i)^{-1} \right) / (N - k) .$$

In some cases, this will give a nonpositive semidefinite estimator. To guarantee a positive semidefinite estimate, we will do the following : Let λ^* be the smallest root of

$$\left| S_{bb} - \lambda \sum_{i=1}^{N} (1 - a_i^t (\tilde{A}^t \tilde{A})^{-1} a_i)(B_i^t \hat{C}_i^{-1} B_i)^{-1} \right| = 0 .$$

If $\lambda^* < 1$, then we will use the modified estimator

$$\hat{\Psi} = \left(S_{bb} - \lambda^* \sum_{i=1}^{N} (1 - a_i^t (\tilde{A}^t \tilde{A})^{-1} a_i)(B_i^t \hat{C}_i^{-1} B_i)^{-1} \right) / (N-k) .$$

This type of modification has been described in Bock and Peterson [18] and Efron and Morris[19] . As mentioned by Vonesh [20], the need to make this adjustment is suggestive of some type of model misspecification, usually in designating which parameters are random effects.

Step 4. Updating the parameter estimates. Since we now have an estimate of Ψ, we can update the estimates of α and b_i using the marginal distribution of $\hat{\phi}_i$. The population parameters are obtained by minimizing

$$Q_N' = \sum_{i=1}^{N} (\hat{\phi}_i - A_i\alpha)^t(\hat{C}_i + B_i\hat{\Psi}B_i^t)^{-1}(\hat{\phi}_i - A_i\alpha)$$

to get the new estimates

$$\hat{\alpha}' = \left(\sum_{i=1}^{N} A_i^t(\hat{C}_i + B_i\hat{\Psi}B_i^t)^{-1}A_i \right)^{-1} \sum_{i=1}^{N} A_i^t(\hat{C}_i + B_i\hat{\Psi}B_i^t)^{-1}\hat{\phi}_i .$$

The usual estimates of the b_i for known Ψ are empirical Bayes, as in Laird and Ware [2], for example. We can substitute our estimate for Ψ from step 2, obtaining

$$\hat{b}_i = \hat{\Psi}B_i^t(\hat{C}_i + B_i\hat{\Psi}B_i^t)^{-1}\,\tilde{e}_i .$$

We can now stop the procedure, as Vonesh and Carter do, and obtain non-iterative solutions. An alternative would be to update the residuals, and repeat steps 2 through 4 in an iterative fashion until the estimate of ψ converges. In practice, this will probably produce little change in the estimates.

After we have obtained final estimates, we will estimate the individual's true parameter vector as

$$\hat{\phi}_i' = A_i\hat{\alpha} + B_i\hat{b}_i .$$

The updated estimate can also be expressed as

$$\hat{\phi}_i' = A_i\hat{\alpha} + W_i\hat{\phi}_i - W_iA_i\hat{\alpha} = W_i\hat{\phi}_i + (I - W_i)A_i\hat{\alpha}$$

where

$$W_i = B_i(B_i^t\hat{C}_i^{-1}B_i)B_i^t\hat{C}_i^{-1} .$$

In this form, it is easy to see that the estimate is a weighted combination of the within individual estimate and the population estimate.

Our estimates of the population parameters will have estimated variance-covariance matrix

$$\hat{Var}(\hat{\alpha}) = \left[\sum_{i=1}^{N} A_i^t(\hat{C}_i + B_i\hat{\Psi}B_i^t)^{-1}A_i \right]^{-1} .$$

Vonesh and Carter [20] discuss a similar global two stage approach. They considered the case where the first stage uses nonlinear least squares and the second stage EGLS. They prove for the special case of a single group that for the estimates of the fixed effects to be consistent, then both N and the minimum n_i must go to infinity. Our procedure may have similar asymptotic properties, especially with respect to N. The requirement of the minimum n_i being large is less certain in our procedure. We improved the first stage, where we replaced the biased nonlinear

least squares estimates with the essentially unbiased Adjusted Composite Periodogram estimators. More research needs to be done into the asymptotic properties of our procedure.

SIMULATION STUDY

A simulation was conducted to verify that the Adjusted Composite Periodogram estimation procedure and the Two Stage EGLS methodology give reasonable results when used together in analyzing the Mixed Effects Cosinor Model. The basic experiment to be simulated consisted of data series of 120 points for each of 10 individuals. First, 10 sets of parameter values were randomly generated as

$$\begin{pmatrix} \alpha_i \\ \beta_i \\ \omega_i \end{pmatrix} \sim N_3 \left(\begin{bmatrix} -0.60 \\ -0.20 \\ 0.2618 \end{bmatrix}, \begin{bmatrix} 0.0300 & -0.0090 & 0.0005 \\ -0.0090 & 0.0100 & -0.0003 \\ 0.0005 & -0.0003 & 0.0001 \end{bmatrix} \right).$$

Using these parameter values, the data were then randomly generated from the Cosinor model

$$y_i = \alpha_i \cos(\omega_i t) + \beta_i \sin(\omega_i t) + e_{it} \qquad t = 1, 2, \ldots 120$$

with $e_{it} \sim N(0, 0.1)$. These data were then analyzed by the Two Stage EGLS methodology with the Adjusted Composite Periodogram estimation procedure used in the first stage. The basic experiment and analysis was replicated in this fashion 1000 times. Table 1 gives the results of this simulation. They indicate that the Mixed Effects Cosinor Methodology performs very well under the conditions of this simulation.

Table 1. Mean Parameter Estimates from 1000 Simulated Experiments

Parameter	True Value	Mean	Std. Dev.
α	–0.6000	-0.60467	0.05846
β	–0.2000	-0.20258	0.04079
ω	0.2618	0.26173	0.00321
ψ_{11}	0.0300	0.03067	0.01632
ψ_{12}	-0.0090	-0.00919	0.00832
ψ_{13}	0.0005	0.00047	0.00065
ψ_{22}	0.0100	0.01053	0.00695
ψ_{23}	-0.0003	-0.00028	0.00045
ψ_{33}	0.0001	0.00010	0.00005

DATA EXAMPLE - PROLACTIN LEVELS IN SHEEP

In this data example, we examine the seasonal rhythmicity of prolactin levels in sheep, and the effect of environment on these levels. The study consists of 10 sheep, five of which were housed under normal outdoor conditions and five of which were housed under controlled, indoor conditions. Prolactin levels (ng/ml) were obtained twice each week for approximately four years from early 1983 through early 1987. We use a subset of these data, consisting of 279 points from July 27, 1985 to March 27, 1987. This range of data was chosen to allow the sheep which were moved indoors to acclimate to their new conditions. The data were log transformed.It is of interest to estimate the rhythm parameters for these two groups, and to compare them.

Individual parameter estimates were obtained using the adjusted Composite Periodogram estimators, and the fits were very reasonable overall. We now consider the second stage model. The between animal design matrix is constructed designating each animal to one of two groups. We first analyze the data considering all of the

parameters to be random effects from the same distribution. To do this we choose $B_i = I_3$. In this setup, the matrix Ψ has 6 unique elements to be estimated. The estimates of the fixed effects and Ψ are given in Table 2. We also are interested in whether the fixed effects are different for the two groups. This was examined by testing the following hypothesis $H_1: \alpha_1 = \alpha_2$; $H_2: \beta_1 = \beta_2$; and $H_3: \omega_1 = \omega_2$. The hypothesis H_1: and H_2: were rejected by both the Chi-square and *F* approximations, while H_3 was not rejected by either. The frequency estimates correspond to periods of 361.9 and 371.4 days, respectively. These are essentially one year periodicities.

Examination of the individual parameter estimates indicates that in the control, outdoor animals the parameter estimates seem to be less variable than in the experimental animals. To account for this in the model, we allow the variance of the random effects to be different in the different groups. This is accomplished by using $B_i = A_i$. The random effects covariance Ψ will now have the form

$$\Psi = \begin{pmatrix} \Psi_1 & 0 \\ 0 & \Psi_2 \end{pmatrix}$$

and has 12 unique elements to be estimated. The results of this analysis are given in Table 3. The estimates of the fixed effects are essentially the same as before. As before, the hypotheses H_1 and H_2 were rejected by both the Chi-square and *F* approximations, while H_3 was not rejected by either. The most striking result of this analysis is that the variation in the frequency parameter in the experimental animals is about 3000 times the variation in the control animals. The estimate of Ψ had to be adjusted to assure positive definiteness, suggestive of model misspecification.

The extremely small estimate of frequency variation in the control animals suggest that it could be considered a fixed effect, while the much larger variation in the experimental animals indicates we may want to leave it as a random effect. We can accommodate this in our second stage model by appropriate choices of the B_i. We use the following

Table 2.

Population Parameter	Estimate	Std. Error
α_1	0.6875	0.03768
β_1	-0.4748	0.05619
ω_1	0.0608	1.219E-3
α_2	0.3384	0.03718
β_2	0.1101	0.05664
ω_2	0.0592	1.264E-3

$$\hat{\Psi} = \begin{pmatrix} 0.005885 & -0.004259 & -0.000055 \\ -0.004259 & 0.01390 & 0.00024 \\ -0.000055 & 0.00024 & 7.287E\text{-}6 \end{pmatrix}$$

Table 3

Population Parameter	Estimate	Std. Error
α_1	0.6868	0.03352
β_1	-0.4722	0.03613
ω_1	0.0608	1.643E-4
α_2	0.3344	0.02278
β_2	0.1110	0.04997
ω_2	0.0592	1.281-3

$$\Psi_1 = \begin{pmatrix} 0.004981 & -0.001068 & 2.709E\text{-}6 \\ -0.001068 & 0.004693 & -2.797E\text{-}6 \\ 2.709E\text{-}6 & -2.797E\text{-}6 & 2.574E\text{-}9 \end{pmatrix}$$

$$\Psi_2 = \begin{pmatrix} 0.001624 & -0.003088 & -0.000059 \\ -0.003088 & 0.01051 & 0.00026 \\ -0.000059 & 0.00026 & 7.559E\text{-}6 \end{pmatrix}$$

$$B_i = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix} \qquad i = 1,2,3,4,5 \qquad B_i = \begin{pmatrix} 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix} \qquad i = 6,7,8,9,10 .$$

Using these choices of B_i results in

$$\phi_i = \begin{pmatrix} \alpha_1 + b_{1i} \\ \beta_1 + b_{2i} \\ \omega_1 \end{pmatrix} \qquad i = 1,2,3,4,5 \qquad \phi_i = \begin{pmatrix} \alpha_2 + b_{3i} \\ \beta_2 + b_{4i} \\ \omega_1 + b_{5i} \end{pmatrix} \qquad i = 6,7,8,9,10 .$$

The data were reanalyzed using this model. With this specification, we don't have to adjust the estimate of Ψ. The results are given in Table 4.

These analyses seem to indicate that environment influences the prolactin cycle in sheep. Animals subjected to outdoor conditions have stronger evidence of a regular cycle, as indicated by the larger amplitude in the Cosinor model. They are also more closely entrained to the seasons, with their frequency well modelled as a fixed effect of approximately a one year period. Animals brought indoors show evidence of decreased amplitude and in the absence of the seasonal influence,
more widely varying frequencies. For these animals, the frequency is adequately modelled as a random effect.

SUMMARY

In this paper, a new procedure for analyzing nonlinear mixed effects models is proposed. It is a two-stage procedure, requiring estimation of individual parameter vectors and variances in a separate first stage. These estimates are then used as input to the second stage. The second stage is a four-step procedure similar to the procedure of Vonesh and Carter [4]. It utilizes generalized least squares for estimation of the fixed effects and the method of moments to estimate the variance of the random effects. This overall procedure allows us flexibility in the estimation procedure of the first stage. This is very important for the Mixed Effects Cosinor Model, since we desire to use the Adjusted Composite Periodogram estimators in the first stage.

Table 4

Population Parameter	Estimate	Std. Error
α_1	0.6867	0.03336
β_1	-0.4721	0.03761
ω_1	0.0608	1.612E-4
α_2	0.3329	0.02116
β_2	0.1109	0.04846
ω_2	0.0592	1.261E-3

$$\Psi_1 = \begin{pmatrix} 0.004365 & -0.001161 \\ -0.001161 & 0.005235 \end{pmatrix}$$

$$\Psi_2 = \begin{pmatrix} 0.001290 & -0.002922 & -0.000055 \\ -0.002922 & 0.009815 & 0.00025 \\ -0.000055 & 0.00025 & 7.328E-6 \end{pmatrix}$$

REFERENCES

1. Lindstrom, M.J. and Bates, D.M. "Nonlinear mixed effects models for repeated measures data." Biometrics vol. 46, pp. 673-687, 1990.

2. Laird, N.M. and Ware, J.H. "Random effects models for longitudinal data." Biometrics , vol. 38, pp. 963-974, 1982.

3. Jennrich, R.I. and Schluchter, M.D. " Unbalanced repeated-measures models with structured covariance matrices." Biometrics , vol. 42, pp. 805-820, 1986.

4. Vonesh, E.F. and Carter, R.L. "Efficient inference for random-coefficient regression models with unbalanced data." Biometrics , vol. 43, pp. 617-628, 1987.

5. Steimer, J.L., Mallet, A., Golmard, J.L. and Boisvieux, J.F. "Alternative approaches to estimation of population pharmacokinetic parameters : Comparison with a nonlinear mixed-effect model." Drug Metabolism Reviews , vol. 15, pp. 265-292, 1984.

6. Racine-Poon, A. "A Bayesian approach to nonlinear random-effects models." Biometrics , vol. 41, pp. 1015-1023, 1985.

7. Sheiner, L.B. and Beal, S.L. "NONMEM User's Guide - Part 1: Users Basic Guide." Division of Clinical Pharmacology, University of California, San Francisco, 1979.

8. Vonesh, E.F. and Carter, R.L. "Mixed-effects nonlinear regression for unbalanced repeated measures." Biometrics, vol. 48, pp. 1-18, 1992.

9. Hallberg, F., Tong, Y.L., and Johnson, E.A. "Circadian system phase - an aspect of temporal morphology; Procedures and illustrative examples." In The Cellular Aspects of Biorhythms , pp. 20-48, Springer-Verlang, Berlin, 1965.

10. Halberg, F., Johnson, E.A., Nelson, W., Runge, W., Sothern, R. "Autorhythmometry - Procedures for physiologic self-measurements and their analysis." Physiology Teacher, vol. 1, pp. 1-11, 1972.

11. Nelson, W., Tong, Y.L., Lee, J.K. and Halberg, F. "Methods for Cosinor rhythmometry." Chronobiologia, vol. 6, pp. 305-323, 1979.

12. Bingham, C., Arbogast, B., Cornelissen, G. Lee, J.K. and Halberg F. " Inferential statistical methods for estimating and comparing Cosinor parameters." Chronobiologia ,vol. 9, pp. 397-439, 1982.

13. Monk, T.H. and Fort, A. " 'COSINA' : A Cosine curve fitting program suitable for small computers." International Journal of Chronobiology , vol. 8, pp. 193-224, 1983.

14. Vokac, M. "A comprehensive system of Cosinor treatment programs written for the Apple II microcomputer." Chronobiology International , vol. 1, pp. 87-92, 1984.

15. Rice, J.A. and Rosenblatt, M. "On frequency estimation." Biometrika , vol. 75, pp. 477-484, 1988.

16. Sollberger, A. "Basic considerations in the study of rhythms." Annals of the New York Academy of Sciences vol. 98, pp. 757-774, 1962.

17. Weaver, R.J. "Analysis of rhythmic biological data : The Mixed Effects Cosinor Model." Unpublished Ph.D. thesis, University of Michigan, Ann Arbor, 1993.

18. Bock, R.D. and Petersen, A.C. "A multivariate correction for attenuation." Biometrika , vol. 62, pp. 673-678, 1975.

19. Efron, B. and Morris, C. "Empirical Bayes on vector observations: An extension of Stein's method." Biometrika, vol. 59, pp. 335-347, 1972.

20. Vonesh, E.F. "Nonlinear models for the analysis of longitudinal data." Unpublished manuscript, 1990.

PRIORITY SELECTION TO MINIMIZE EXPECTED LOSS

Dr. Douglas H. Frank
Indiana University of Pennsylvania
Department of Mathematics
208 Stright Hall
Indiana, PA 15705-1072

Ann E. M. Brodeen
U.S. Army Research Laboratory
Information Science and Technology Directorate
Aberdeen Proving Ground, MD 21005-5067

ABSTRACT

Target values are assessments keyed to the enemy's perception of the function of its assets, assets the enemy threat commander requires for the successful completion of his mission. This report discusses the assignment of target values, as an aid to target selection for engagement, based on stochastic optimization. In particular, we determine values that minimize expected damage to the friendly unit when the enemy targets are engaged in order of decreasing value. This approach has three advantages. First, it is based on a realistic tactical scenario in which k enemy targets are engaged, one by one, by friendly fire until all are destroyed, and the enemy independently returns fire on the friendly forces. Second, it is mathematically amenable and allows us to derive globally optimal results. Third, addressing the psychological and political exigencies to reduce fratricide, it enables us to determine the impact on a target's value given some probability it is actually a friendly target. Though the basis of this research is fire support targeting, it has potential application in any scheduling or rationing framework (i.e., multimedia networks; allocation of medical resources; scheduling vehicle maintenance).

INTRODUCTION

In our previous work, we defined target value to maximize the damage inflicted on an array of enemy targets.[1,2] We call that the "Damage Inflicted" model. This gave some interesting, but not completely satisfying, results. In this paper, we consider an approach to minimize the damage inflicted by the enemy. We call this the "Damage Received" model. This seems to be a better model for at least three reasons:

(1) it gives exact optimal results,
(2) it allows us to consider other factors such as friendly fire damage, and
(3) the model reflects a more realistic battle assessment.

We will illustrate these points for a simple discrete shot battle.

THE MODEL

We consider a battle in which a friendly battery engages k enemy targets. The strategy is to engage a single enemy target until it is removed before firing upon the next target in the ordering. The battery itself is not fired upon. In the discrete shot battle, each enemy target fires one shot against the remaining friendly forces in one time unit. Hits are independent; each hit results in one unit of damage. The battle continues until all k enemy targets are removed. The expected total damage to the friendly forces is the aggregate of the damage inflicted by each of the k enemy targets.

Approved for public release; distribution is unlimited.

Consider the following parameters for each enemy target i = 1, 2, . . . , k:

k = number of enemy targets

p_i = single shot probability of removing target i

r_i = single shot probability of target i inflicting damage

f_i = probability of target i being friendly.

THE DISCRETE BATTLE

DEFINITIONS

- D_{ij} is the damage inflicted by target j while target i is being engaged, where $i \leq j$ and i, j are enemy targets.
- D is the total damage inflicted by the enemy during the battle.
- D_T is the damage inflicted by the enemy during the battle if enemy targets i and i+1 are interchanged in the target engagement ordering.

We shall establish these definitions in the following Lemmas.

LEMMA 1

The expected value of D_{ij} is

$$E\left[D_{ij}\right] = \frac{r_j}{p_i}.$$

The proof is as follows. Let N_i be the number of shots until enemy target i is destroyed. Then,

$$E\left[D_{ij}\right] = E\left[D_{ij} \mid N_i\right] = E\left[N_i r_j\right] = \frac{r_j}{p_i},$$

since N_i is geometrically distributed.

Since D is the cumulative damage inflicted by the enemy during the battle, the next result follows immediately.

If the enemy targets are engaged in the order 1, 2, . . . , k, the expected total damage incurred by the friendly forces is

$$E[D] = \sum_{i=1}^{k} \sum_{j=i}^{k} \frac{r_j}{p_i} = \sum_{i=1}^{k} \sum_{j=i}^{k} D_{ij}.$$

This pertains to any order of engagement, especially one that is not determined with the aid of a value algorithm.

To obtain an optimal target value ordering based on the parameters of interest, we need to examine the effect on the damage incurred by the friendly forces if the target engagement ordering is interchanged. The following is a key result.

LEMMA 2

If adjacent enemy targets i and i+1 are interchanged, where i = 1, 2, . . . , k-1, the expected total damage increases; that is,

$$E\left[D_T\right] > E\left[D\right] \quad \text{iff} \quad p_i \cdot r_i > p_{i+1} \cdot r_{i+1}.$$

The proof is as follows. The battle consists of i = 1, 2, . . . , k segments. In each segment, one of the enemy targets present is engaged. We assume the damage inflicted by the enemy during a segment is independent of both the target engagement ordering and the damage inflicted during any other segment of the battle. That is, the expected damage in segment i is unchanged if any pair of targets, other than the pair (i, i+1), is interchanged.

If adjacent targets i and i+1 are transposed, where i = 1, 2, . . . , k–1, consider the expected damage over segments i and i+1 for, first, the transposed target engagement ordering and, second, the target engagement ordering in the natural order. The difference between the expected damage for each condition is

$$E\left[D_T\right] - E\left[D\right] = \left(\frac{r_{i+1} + r_i}{p_{i+1}} + \frac{r_i}{p_i}\right) - \left(\frac{r_i + r_{i+1}}{p_i} + \frac{r_{i+1}}{p_{i+1}}\right)$$

$$= \frac{r_i}{p_{i+1}} - \frac{r_{i+1}}{p_i}.$$

Therefore,

$$E\left[D_T\right] - E\left[D\right] > 0 \quad \text{iff} \quad \frac{r_i}{p_{i+1}} - \frac{r_{i+1}}{p_i} > 0,$$

and

$$E\left[D_T\right] > E\left[D\right] \quad \text{iff} \quad p_i \cdot r_i > p_{i+1} \cdot r_{i+1}.$$

For the discrete battle, if we define p_i as the vulnerability of target i and r_i as the threat of target i, the value may be expressed as the product of the vulnerability and the threat.

DEFINITION

The value of enemy target i is

$$V_i = p_i \cdot r_i .$$

THEOREM 1

In a discrete battle, if the enemy targets are engaged in order of decreasing value, the expected total damage to the friendly forces is a minimum.

Consider any pair of targets in which the left-hand target has a smaller value. Lemma 2 asserts that the expected total damage will decrease when the targets are transposed. This process may be continued until the target array is arranged in order of decreasing value.

FRIENDLY FIRE

Close to the surface of battle planning processes, the aspect of friendly fire casualties influences decisions and operations orders. To introduce this factor into the target value algorithm, we define a new parameter, f_i, the probability that target i is friendly. We redefine the threat of target i as $r_i\ (1 - f_i)$ and determine the impact of this parameter on a target's value with respect to reducing expected damage.

Optimal results are easily obtained for the discrete battle. The approach considers a hit on a friendly target to be equivalent to one unit of damage; it also assumes the presumed friendly target inflicts no return damage on the friendly forces. Only minimal changes to the previous section are required to introduce the friendly fire parameter.

LEMMA 3

In a discrete battle with possibly friendly targets, where $i = 1, 2, \ldots, k$, the expected damage is

$$E\left[D_{ij}\right] = \begin{cases} \dfrac{\left(1 - f_j\right) r_j}{p_i}, & \text{if } i < j \\ \dfrac{\left(1 - f_j\right) r_j}{p_i} + f_j, & \text{if } i = j. \end{cases}$$

The proof is as follows. To examine the possible damage to the friendly forces from target j while target i is being engaged, let F_j be the event that target j is actually friendly, then

$$E\left[D_{ij}\right] = E\left[D_{ij} \mid F_j\right] \cdot f_j + E\left[D_{ij} \mid \bar{F}_j\right] \cdot (1 - f_j).$$

Obviously, $E\left[D_{ij} \mid F_j\right] = 0$, if $i < j$ (recall that the friendly target is not returning fire; therefore, there will be no expected damage), and $E\left[D_{ij} \mid F_j\right] = 1$, if $i = j$ (i.e., a friendly target is destroyed).

$E\left[D_{ij} \mid \bar{F}_j\right]$ is equivalent to $E\left[D_{ij}\right]$ of Lemma 1. Using the arguments of the previous section, we obtain the optimal results.

DEFINITION

The value of enemy target i is

$$V_i = \left(p_i \cdot r_i\right)\left(1 - f_i\right).$$

THEOREM 2

Given a discrete shot battle with possibly one or more friendly targets in the strike zone, if the targets are engaged in order of decreasing value, the expected total damage to the friendly forces is a minimum.

THE DISCRETE BATTLE WITH FINITE AMMUNITION SUPPLY

Does the available ammunition load affect the target engagement ordering? In particular, if the battle terminates after N shots have been fired (or all k targets have been removed), is there a target engagement ordering that minimizes expected damage? The answer is not immediately obvious and the algebra much more painful, but the key result parallels that for the discrete battle.

Since the ammunition supply is finite, the number of shots to remove a target is a finite random variable; hence, the proof of Lemma 1 does not apply.

Assume the same discrete battle in which the friendly and enemy forces have N rounds of available ammunition each. The battle will end when one side has expended all N rounds.

Denote the portion of the battle in which enemy target i is engaged as the i^{th} segment. Let T_i be the length of segment i (i.e., the number of shots until enemy target i is removed or the ammunition supply is exhausted). T_i is a geometrically distributed random variable with success probability p_i and censored at N.

LEMMA 4

Let T_1 be the number of shots to remove target 1, with success probability p_1. It can be shown that $P(T_1 \geq n) = q_1^{n-1}$, thus

$$E\left[T_1\right] = \frac{1 - q_1^N}{p_1}, \quad \text{where } q_1 = 1 - p_1.$$

To verify, we make use of results from the theory of expected values of positive discrete random variables and the theory of finite geometric series. For $n = 1, 2, \ldots, N$

$$E\left[T_1\right] = \sum_{n=1}^{N} P(T_1 \geq n) = \sum_{n=1}^{N} q_1^{n-1}$$

$$= \frac{1 - q_1^N}{(1 - q_1)}.$$

Let T_2 be the number of shots to remove target 2, with success probability p_2, then

$$E\left[T_2\right] = \begin{cases} \dfrac{1}{p_2} - \dfrac{p_1 q_2^N - p_2 q_1^N}{p_2\left(q_2 - q_1\right)}, & \text{if } p_1 \neq p_2 \\ \dfrac{1}{p} - (N-1)\, q^{N-1} - \dfrac{q^{N-1}}{p}, & \text{if } p_1 = p_2 = p. \end{cases}$$

The proof applies the preceding results with a fixed number of shots to remove target 1, which we defined as T_1. Looking at segment 2 with an ammunition supply reduced to $N - T_1$,

$$E\left[T_2 \mid T_1\right] = \frac{1 - q_2^{N-T_1}}{p_2}.$$

Thus,

$$E\left[T_2\right] = \frac{1}{p_2} - \frac{1}{p_2} E\left[q_2^{N-T_1}\right],$$

where

$$E\left[q_2^{N-T_1}\right] = \sum_{n=1}^{N-T_1} p_1 q_2^{n-1} \left(\frac{q_1}{q_2}\right)^{n-1} + q_1^{n-1}$$

$$= \frac{p_1 q_2^N - p_2 q_1^N}{q_2 - q_1}, \qquad \text{if } p_1 \neq p_2$$

and

$$= (N-1)\, p\, q^{N-1} + q^{N-1}, \qquad \text{if } p_1 = p_2 = p.$$

From Lemma 4, we can state the following theorem.

THEOREM 3

The expected damage in a discrete battle with two enemy targets and finite ammunition N is

$$E[D] = (r_1 + r_2)\, E[T_1] + r_2\, E[T_2].$$

From this it can be shown, in a fashion similar to that for the discrete battle with infinite ammunition, that

$$E[D_T] > E[D] \text{ iff } p_1 \cdot r_1 > p_2 \cdot r_2.$$

We can apply the methodology behind Lemma 4 to derive E $[T_3]$, E $[T_4]$, and so forth, but the results become increasingly complex. This was, in fact, carried out for the three-target case; computer calculations of expected damage for the six permutations led to the following surprising definition and theorem.

DEFINITION

The value of enemy target i is

$$V_i = p_i \cdot r_i.$$

THEOREM 4

For any finite ammunition supply N, the ordering of targets by their value V_i produces a minimum expected total damage to the friendly forces over all k! enemy target engagement orderings.

SIMULATIONS

This paper discusses absolute optimal results for the discrete battles described within. Even though the battles are somewhat simplistic, they do provide some indication of how to rank a target set. A mathematical analysis of a more complex scenario would be too complicated to yield simple target values.

To verify the target value algorithms in a more realistic setting, we studied the models by means of simulation. The battle consisted of four enemy targets, three friendly targets, and one friendly battery. Each enemy target randomly fired at one of the three friendly targets, and each friendly target randomly fired at one of the enemy targets. The battery fired at the enemy in some specified order. The enemy did not engage the battery until all three friendly targets were removed. The battle ended when either side lost all four of its targets.

The battle was simulated 5,000 times for each of the 24 enemy target engagement orderings. We simulated two models: discrete battle with infinite ammunition and discrete battle with finite ammunition. For each case, we calculated the rank correlation to examine the degree of similarity between the theoretical expected damage from the mathematical model and the actual expected damage over the 5,000 simulated battles. The results were impressive (see Table 1).

The theoretical target value models were designed to minimize the damage inflicted by the enemy on the friendly forces. In the more realistic simulated battle scenario, we also observed the frequency with which all enemy targets were removed, an event appropriately defined as victory. Certainly, we would expect the minimizing of enemy-induced damage to have an indirect effect on increasing the friendly forces' chance of victory. It was not surprising to observe a significant negative correlation between rank orderings to minimize loss and the number of victories in the simulated battles (see Table 2).

Table 1. Rank Correlation Between Theoretical and Actual Damage

Model	Correlation
Discrete With Infinite Ammo	0.929
Discrete With Finite Ammo (4 rounds)	0.907

Table 2. Rank Correlation Between Loss and Victory

Model	Correlation
Discrete With Infinite Ammo	–0.928
Discrete With Finite Ammo (4 rounds)	–0.859

CONCLUSIONS

The models presented have several strengths.

- They are somewhat realistic.
- They allow us to derive simple, intuitive values for targets.
- They consider the influence of intelligence information (i.e., the possibility of a friendly target in the firing sector) on the values.

The simulations support the theoretical results. Thus, the product of vulnerability and threat seems to produce a good value for ranking targets to produce optimal results. Further consideration should be given to constructing better models to assess friendly fire. We also need to develop more sophisticated battle simulations to validate the target engagement orderings.

This applied research in optimal target value assessment algorithms may be applied to operations other than war (OOTW) (e.g., rationing medical care in a trauma situation, scheduling vehicle maintenance).

REFERENCES

1. Frank, D. H., and A. E. M. Brodeen. "Target Prioritization to Optimize Expected Utility For a Simple Battle Scenario." BRL-MR-3945, U.S. Army Ballistic Research Laboratory, Aberdeen Proving Ground, MD, October 1991.

2. Frank, D. H., and A. E. M. Brodeen. "Target Prioritization to Optimize Expected Utility For a Random Battle Scenario." BRL-MR-3998, U.S. Army Ballistic Research Laboratory, Aberdeen Proving Ground, MD, September 1992.

BIBLIOGRAPHY

1. Bhattacharya, R. N., and E. C. Waymine. Stochastic Processes With Applications. New York: John Wiley & Sons, Inc., 1990.

2. Brodeen, A. E. M., and W. A. Winner. "Classification Tree Methodology: Another Approach to the Allocation and Distribution of 155-mm Howitzer Fire." BRL-MR-3682, U.S. Army Ballistic Research Laboratory, Aberdeen Proving Ground, MD, July 1988.

3. Dougherty, W., and R. Kaste. "Knowledge Acquisition Survey and Analysis: Firepower Control Experiment Part 11 of 12." BRL-MR-3731, U.S. Army Ballistic Research Laboratory, Aberdeen Proving Ground, MD, December 1988.

4. Fishburn, P. C. Utility Theory for Decision Making. New York: John Wiley & Sons, Inc., 1970.

5. Gellert, W., H. Küstner, M. Hellwich, and H. Kästner, editors. The VNR Concise Encyclopedia of Mathematics. New York: Van Nostrand Reinhold Company, 1977.

6. Kemeny, J. G., and J. L. Snell. Finite Markov Chains. Princeton: D. Van Nostrand Company, Inc., 1960.

7. Olkin, I., L. J. Gleser, and C. Derman. Probability Models and Applications. New York: Macmillan Publishing Company, Inc., 1980.

8. Pratt, J. W., H. Raiffa, and R. O. Schlaifer. Introduction to Statistical Decision Theory. Preliminary ed., New York: McGraw-Hill, 1965.

9. U.S. Army Field Artillery School. "Targeting and Target Value Analysis." Coordinating Draft, FC 6-20-2, Fort Sill, OK, October 1984.

10. von Neumann, J., and O. Morgenstern. Theory of Games and Economic Behavior. 2nd ed., Princeton, NJ: Princeton University Press, 1947.

METHODOLOGY FOR THE CURVE FITTING OF NONLINEAR RIDE CURVES

Andrew W. Harrell
U.S. Army Engineer Waterways Experiment Station
Vicksburg, MS 39180-6199

ABSTRACT

This paper discusses the application of a non-linear regression technique for describing the relationship between vehicle ride performance and surface roughness. Curves considered for the best-fit come from a two parameter linear envelope of hyperbolas whose asymptotes are the vertical and horizontal axes. The non-linear curve fitting method utilizes the singular-value decomposition of the design matrix of the curve fitting problem. This matrix evaluates the two members of the envelope of functions at the data points. A procedure was developed to use a combination simple search of the parameter space along with a varied solution of the Marquardt minimization procedure. The primary source of data for fitting the curves comes from a series of experimental tests of military vehicles conducted over the last 25 years in various locations. The results of the fitting,in terms of sums of squares of residuals, were examined as a represention of simulations which used a vehicle dynamics model. These results were compared to those obtained with another curve fitting method in order to validate the procedure. This other method used a simple search of the parameter space initiated by a three point interpolation formula. Curves for vehicle test data were calculated and compared with the curves which had been drawn manually.

INTRODUCTION

PURPOSE. To develop a non-linear regression methodology. This methodology should be applicable to accuratedly and consistently representing surface roughness versus ride-limited speed relationships. It should allow for extrapolation beyond the data ranges. The curve produced can then be used in the Nato Reference Mobility Model (NRMM) vehicle speed prediction program and also to determine the effectiveness of VEHDYN2 as a ride simulator (Ahlvin 1992 [1], Creighton 1986 [2]).

SCOPE. This paper will examine several numerical statistical methods each of which involves computing estimates for several undetermined coefficients of basis function approximations. The standard non-linear least-squared approach and a singular-valued matrix approach are examined first. Then the more general Marquardt

method, which allows the undetermined coefficients to enter in a non-linear way with respect to the basis approximation functions, is examined. Also, a simple coefficient search method was employed for comparison with the first two methodologies. Graphs displaying the results of the fitting are computed, and tables measuring the effect of varying tire pressure on the residuals to the fit are calculated. A full discussion of the different methods of how ride curves are determined and the non-linear regression methods which are applicable is contained in the pending report by Harrell [5], "Methodology for the Curve Fitting of Nonlinear Ride Curves," listed in the bibliography.

METHODOLOGY TO FIT NONLINEAR CURVES TO THE SURFACE ROUGHNESS/RIDE LIMITED SPEED RELATIONS[1]

The restrictions on a ride curve that helps it to be determined are the location of the asymptotes. These asymptotes are determined in terms of test data by plotting the ride limited speed on the vertical and the surface roughness on the horizontal axis. The vertical asymptote must be either the y-axis or right of the y-axis because it is postulated that it is possible for any vehicle to approach infinite ride-limited speed on a completely smooth surface. The vertical asymptote must also stay to the left of the first data point because that data point proves there is a limit on speed at that surface roughness. The horizontal asymptote is above the x-axis because it is postulated that it is theoretically possible for a vehicle to cross any surface as long as it goes slow enough.

Two different approaches were taken to find a solution for this hyperbola. In the first approach, a method was taken that would search all possible coefficients for the best fitting curve. The second approach manipulated existing MATHCAD functions so they would give hyperbolas which met the requirements mentioned above. Both methods gave satisfactory ride curves.

DIRECT SEARCH METHOD. The searching method was first employed in a Fortran program which attempted to find the best coefficients (A,B,C) for the equation below.

---

[1] RMS is an acronym used in the characterization of the surface roughness of terrain. It means root mean squared. It is determined by first detrending the surface elevation measurements taken at one foot intervals in a terrain profile and then computing the ordinary square root of the variances of the measurements from the detrended value. Ride limiting speed represents the speed at which vehicle vibrations at the driver seat reach absorbed power limits of 6-watts.

$$y = \frac{A}{x + C} + B$$

The original Fortran program which, was written by Richard Ahlvin in 1982, varied A and B between 0 and 20, and C between 0 and 0.6. These values were determined by examining the range of the data from the experiments and trying different values for the coefficients by trial and error. The program computed the sum of squares of the residuals for each possible combination and the lowest was chosen as the best fit. A flowchart for this Fortran program is shown in Figure 1.

The Fortran program was then transferred to a MATHCAD sheet by Mr. Robert Demillio in order to graph the results.

MARQUARDT METHOD AND SINGULAR VALUED DECOMPOSITION METHOD. The other method of plotting ride curves used MATHCAD's built-in function which performs the Marquardt method to fit a curve to a series of data points. In order for this function to be useful in generating ride curves, there had to be a way to restrict the asymptotes. The only way to keep the vertical asymptote within the given constraints was to define the B coefficient as 0, while the A and C coefficients were computed by the Marquardt method. As long as the B coefficient was 0 and the data points were valid, the horizontal asymptote remained above the x-axis. The result with the lowest sum of the squares of the residuals was the best fitting hyperbola.

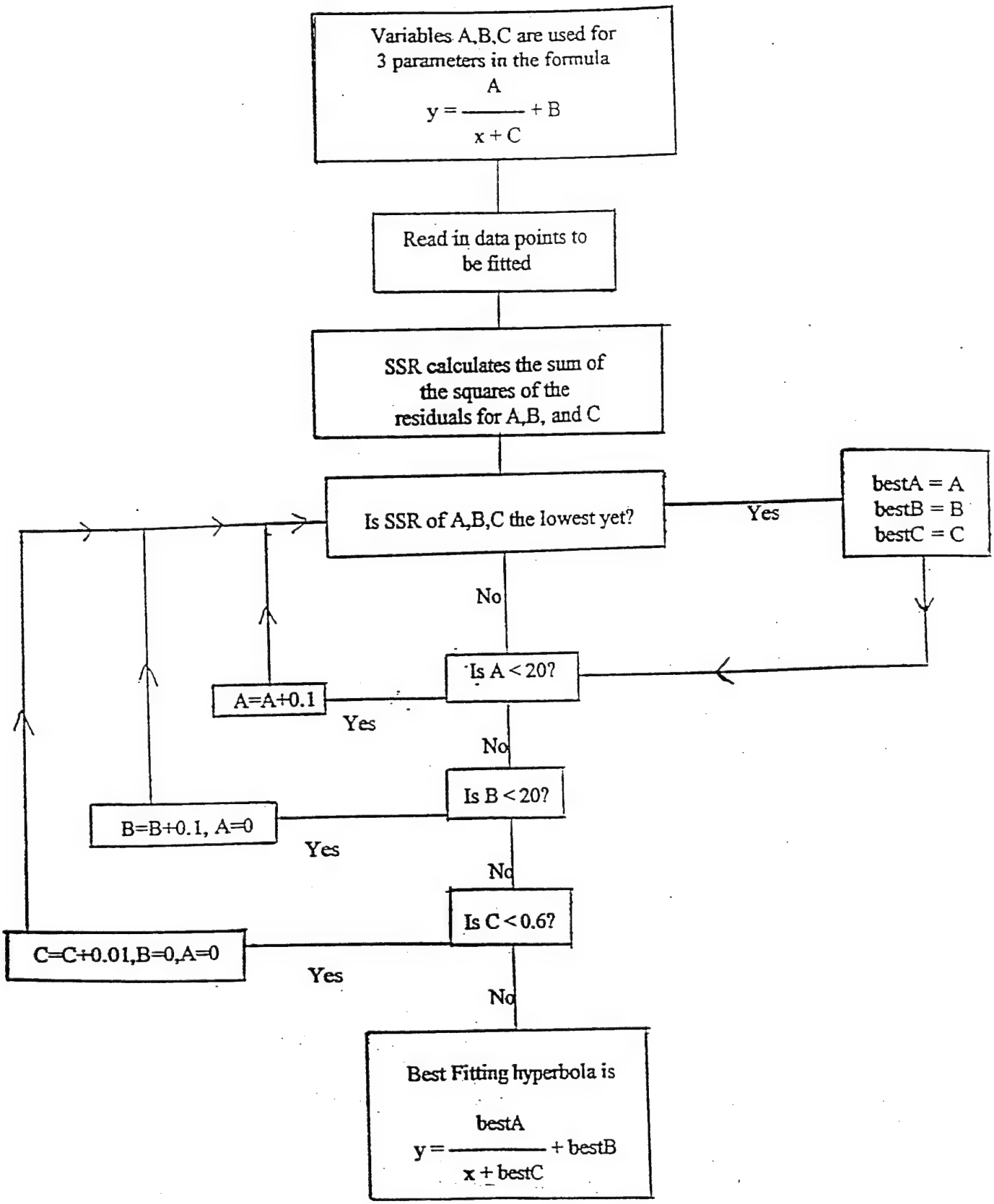


Figure 1 Flowchart of a Fortran program to find the coefficients of the ride curve by minimizing the sum of squares of residuals using a direct search of the coefficient space

For the second method,we follow the approach outlined in Press et al, 1992 [11].

The general form of the model is:

$$y(x) = \sum_{k=1}^{M} a_k * X_k(x)$$

where $X_1(x),...X_M(x)$ are arbitrary fixed functions of x, called the basis functions. For example, the functions could be: $1,x,x^2,....,x^{M-1}$.

In order to generalize the approach to linear least squares fitting, we introduce the function:

$$\chi^2 = \sum_{i=1}^{N} \left[\frac{y_i - \sum_{k=1}^{M} a_k * X_k(x_i)}{\sigma_i}\right]$$

where the symbol σ_I in the denominator is the standard deviation or measurement error of the ith data point.

Taking the derivative of the above expression with respect to all the m parameters a_k , setting it equal to zero, regrouping and renaming the variables in terms of covariance functions yields a set of linear matrix equations (called the normal equations of the fitting problem) to solve

$$\sum_{j=1}^{M} \alpha_{kj} * a_j = \beta_k$$

where

$$\alpha_{kj} = \sum_{i=1}^{N} \frac{X_j(x_i) * X_k(x_i)}{\sigma_i^2}$$

and

$$\beta_k = \sum_{i=1}^{N} \frac{y_i * X_k(x_i)}{\sigma_i^2}$$

These matrix equations can be solved by either Gauss-Jordan elimination procedure or, in the case, when the normal equations are very close to singular, by the singular value decompostion approach.

These functions are now implemented in the MATHCAD spreadsheet, symbolic calcuation program.

As a first approach, we assume that in our problem the hyperbolas of the fitting problem will all have the x and y axis as asymptotes. Therefore an envelope or family of possible fitting curves can be defined as:

$$y(x) = \frac{A}{x} + B$$

where A and B are two parameters to be determined.
This is an example of the previous general non-linear curve fitting problem where $x_1 = 1/x$, $x_2 = 1$, and $a_1 = A$, $a_2 = B$.

The above family of hyperbolas gives a best hyperbolic fit to the data using two arbitrary parameters which are linear coefficients of two basis functions 1/x and 1. If, however, it is desired to consider the effect of translating these curves parallel to the x-axis, it is necessary to use a non-linear fitting scheme.

If we solve the formulas determining the above curves for x, we get:

$$x(y) = \frac{A}{y-B}$$

We can see the effect of translating these curves parallel to the x-axis by adding another parameter C to this family:

$$x(y) = \frac{A}{y-B} + C$$

Solving for y and switching the names of two of the parameters, we get a three parameter family of non-linear curves:

$$y(x;A,B,C) = \frac{A}{x-B} + C$$

To solve this problem we use a modification of the methodology above. We seek to minimize:

$$\chi^2 = \sum_{i=1}^{N} \left[\frac{y_i - y(x;A,B,C)}{\sigma_i}\right]$$

We have the same equation to solve:

$$\sum_{j=1}^{M} \alpha_{kj} * a_j = \beta_k$$

but now:

$$\alpha_{kj} = \frac{1}{2} \frac{\partial^2 \chi^2}{\partial a_k \, \partial a_j}$$

where a_k is one of the parameters: a,b,c,

and

$$\beta_k = -\frac{1}{2}\frac{\partial \chi^2}{\partial a_k}$$

The solution search method of Marquardt, as implemented in the C language in the reference by Press et.al, 1992 [11], can now be used to solve these non-linear equations for the three parameters. This method requires an initial approximate guess of the solution. The two parameters determined by the previous linear fitting approach along with a value of 0 for the third parameter C can be used for this initial approximation to the solution.

In order to test the software, we first tried to fit a family of hyperbolas of this type to a set of points generated by adding random numbers to a given hyperbolic equation. The results show that in this situation both software procedures, the Gauss-Jordan, and the singular value decomposition work approximately equally well. However, because of the more flexible ability of the singular value methodology in all circumstances, we decided to use it for this situation. Figure 2 shows a copy of the MATHCAD sheet used to

Program to compute best fit of a three parameter family of hyperbolas to given data.

$j := 0..2$ $v_j := 1$ v tells which coeffs to fit

$NPTS := 8$ $i := 0..8$

initial values $u_0 := 7.668$ $u_1 := 0$ $u_2 := 5.404$

$m := 2$

$x_i := READ(Rms1)$

$y_i := READ(Ridehy4)$

$s_i := 1$

function and its derivatives

$$f(x,g) := \begin{bmatrix} \frac{g_0}{x - g_1} + g_2 \\ \frac{1}{x - g_1} \\ g_0 \cdot (x - g_1)^{-2} \\ 1 \end{bmatrix}$$

This program computes non-linear fitting coefficients A,B,C (shown as g0,g1,g2 above). It uses the Marquardt method. The program uses as initial guesses the values for A and C and computes using the singular-valued decomposition (SVD) the best least squares fit. It uses an initial value of zero for B.

$r := Mrqmin(x,y,s,u,v,f)$

best coeff. values and the sum of squares of residuals for each fitting coeff.

$$r = \begin{pmatrix} 24.818 & 14.705 \\ -0.977 & 0 \\ 0.918 & 0 \end{pmatrix}$$

$j := 0..20$

$w_j := f\left(\frac{j}{7}, r^{<0>}\right)_0$

Best fitting solution

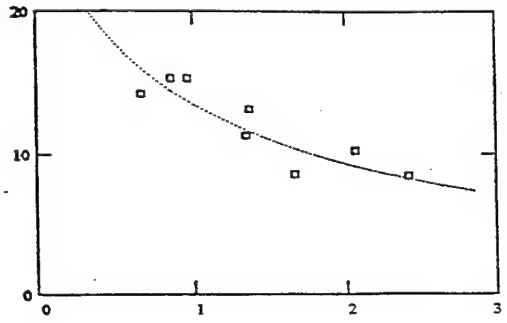
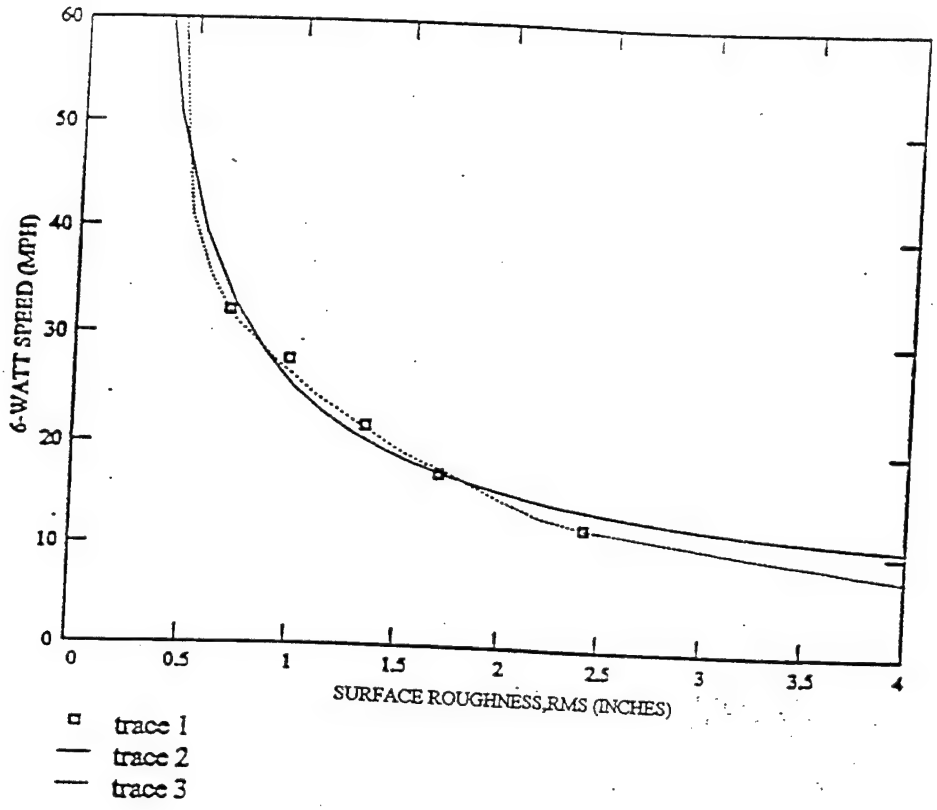


Figure 2. MATHCAD program to compute ride curves.

M1025 HMMWV, 15/29 PSI

Empty



Trace 1 represents the data points collected from 1992 field tests as reported in the Technical Report GL-92.07, *Mobility Performance Tests of the High Mobility Multi-Purpose Wheeled Vehicle with a Central Tire Inflation System and Towed Trailer.* (the pts by themselves are displayed).

Trace 2 represents the 6 watt ride curve for the HMMWV computed by the above Mathcad sheet from the data points in trace 1.

Trace 3 represents the 6 watt ride curve as shown in the report GL-92.07. (the pts. were taken off the curve in the report and connected by straight interpolating lines).

Figure 3. Ride curve for the unloaded 1025 HMMWV

calculate the ride curve. The data points for this program are extracted through a MATHCAD read statement which refers to the corresponding data file. Figure 3 shows the results of applying this procedure to the historical test results, and standard vehicle information on performance.

The results of two of the methods of fitting hyperbolas to the ride curve data are displayed below. Data points from tests of the 10 ton Heavy Expanded Mobility Tactical trucks (HEMMT) were used. The vehicle was tested in both the loaded condition (60,145 lbs) and the unloaded condition (38,018 lbs). Tire pressures are given in pounds per square inch for both the two front and two rear tires. More details about the tests are contained in the report by Schreiner et. al. , 1985 [13]. The Marquardt function was compared with the direct search method. The searching method gave results similar to the Marquardt method which is evidence that the ride curves are accurate. Tables 1 and 2 give a comparison between the two fitting methods.

Table 1

TOTAL CURVE FITTING ERROR MEASURED

SUM OF SQUARES OF RESIDUALS

HEMMT in
Unloaded Condition

Surface Type	**Tire Pressure**	**Search Method**	**Mrqmin Method**
Standard Highway	60/70	128.7	128.6
Cross Country (clay)	35/40	37.2	36.9
Sand	20/30	24.7	24.6
Emergency	15/19	30.0	30.0

Table 2

TOTAL CURVE FITTING ERROR MEASURED IN SUMS OF SQUARES OF RESIDUALS

HEMMT in
Loaded Condition

Surface Type	Tire Pressure	Search Method	Mrqmin Method
Standard Highway	60/70	89.5	89.4
Cross Country (clay)	35/40	57.1	57.0
Sand	20/30	25.7	25.7
Emrgency	15/19	23.7	23.7

After comparing the different methodological approaches it was decided to use the Marquardt method as implemented in the MATHCAD sheet. In order to make sure the Marquardt ride curves were accurate and met all of the restrictions, the ride curves were computed for the HEMTT and compared to the old ride curves. A complete set of the ride curves which have been calculated can be found in the pending report by Harrell [5]. Points for these graphs were taken from the most recent field tests that could be found for each vehicle. To calculate the 6 watt limit, a curve was drawn through the data from the field tests and the speed at 6 watts was taken to be the limit for that RMS value.

CONCLUSIONS AND RECOMMENDATIONS

Both the standard linear least square method and the singular value linear least square method give basically the same results. Both these methods assume that the x-axis and the y-axis are asymptotes of the family of curves. These results can be improved by using the 2-parameter values determined for the linear family of hyperbolas augmented by zero as initial values for a non-linear three parameter family of hyperbolas. For this family, we do not assume that the y-axis is the y-asymptote; only that the y-asymptote is parallel to the y axis. This assumption corresponds with the physical situation that the 0-RMS ride limited speed is not infinity but the vehicle's highest speed. The results of

computing the sum of squares of residuals are displayed in Table 3 below. These residuals will be the vertical distances between the data points with a given x-cordinate and the corresponding point on the fitting curve with that same x-cordinate.

Table 3

COMPARISON OF RESULTS OF FITTING RIDE CURVES TO DATA

Vehicle	Tire Pressure	Tire Deflection	Residuals using, Marquardt method
HEMTT UNLOADED	60/70	2.1 inches	129
	35/40	3.2 inches	37
	20/30	4.3 inches	25
	15/19	4.7 inches	30
HEMTT LOADED	60/70	2.1 inches	89
	35/40	3.2 inches	57
	20/30	4.3 inches	26
	15/19	4.7 inches	24

Another piece of information that is evident from this table is the sensitivity of the effects of changing tire pressure for computation of the NO-GO[2] speed values for this vehicle. If we use the Marquardt method to model the ride curve we see that an upper limit on the total possible error in determination of the ride-limited speed cutoff changes by a factor of 129/30 in the unloaded case and a factor of 89/24 in the loaded case. Possible future work could include the design of a series of experimental ride tests to further validate these conclusions.

[2] Here we mean by NO-GO speed only the ride-limited part of the overall speed prediction program's computation.

ACKNOWLEDGMENTS

The tests described and the resulting data presented herein, unless otherwise noted, were obtained from research conducted under the MILITARY RESEARCH DEVELOPMENT TEST AND EVALUATION PROGRAM of the United States Army Corps of Engineers by the U.S. Army Engineer Waterways Experiment Station. Permission was granted by the Chief of Engineers to publish this information. I wish to thank Robert DeMillio for writing the MATHCAD program to do the direct search approach and also for computing some of the graphs that were used. I also wish to thank Mr. Newell Murphy for reviewing this project, supplying information from previous vehicle tests, and making suggestions about the constraints which should be used for the search of the family of curves and determination of the best fitting equation.

BIBLIOGRAPHY

1. Richard B. Ahlvin and Peter W. Haley, "NATO Reference Mobility Model Edition II, NRMM II User's Guide," Technical Report GL-92-19, December 1992, U.S. Army Waterways Experiment Station, Vicksburg, MS.

2. Daniel C. Creighton, "Revised Vehicle Dynamics Module: User's Guide for Computer Program VEHDYN II," Technical Report SL-86-9, May 1986, US Army Waterways Experiment Station, Vicksburg, MS.

3. Daniel C. Creighton, "A Preprocessor for the Revised Vehicle Dynamics Module: User's Guide for Computer Program PREVDYN2," Technical Report SL-87-9, April 1987, U.S. Army Waterways Experiment Station, Vicksburg, MS.

4. Andrew W. Harrell, "Measures of Effectiveness for Monte Carlo Sensitivity Analyses," Proceedings of the 39th Conference on the Design of Experiments in Army Research Development and Testing, U.S. Army Research Office, ARO Report 94-2, 1994, Durham, NC.

5. Andrew W. Harrell, "Methodology for the Curve Fitting of Nonlinear Ride Curves," U.S. Army Waterways Experiment Station, Vicksburg, MS 39181, pending publication.

6. MathSoft Inc., "MATHCAD 3.1 User's Guide," 1992, Cambridge, MA.

7. MathSoft Inc., "MATHCAD Plus 5.0," 1994, Cambridge, MA.

8. Marquardt, D.W. 1963, "Journal of the Society for Industrial and Applied Mathematics," vol 11, pp. 431-441.

9. Newell R. Murphy, Jr., "Dynamics of Wheeled Vehicles, Report 1

and Report 3," Technical Reports M-68-1 and M-68-3, U.S. Army Waterways Experiment Station, Vicksburg, MS.

10. Newell R. Murphy, Jr., and Richard B. Ahlvin, "AMC-74 Vehicle Dynamics Models," Technical Report M-76-1, Jan 1976, U.S. Army Waterways Experiment Station, Vicksburg, MS.

11. William H. Press, et. al.,"Numerical Recipes in C, The Art of Scientific Computing, 2nd ed.,"Cambridge U. Press, 1992.

12. A.A. Rula and C.J. Nuttall, Jr., "An Analysis of Ground Mobility Models," Technical Report M-71-4, July 1971, U.S. Army Waterways Experiment Station, Vicksburg, MS.

13. Barton G. Schreiner, Dennis W. Moore, and Keafur Grimes, "Mobility Assessment of the Heavy Expanded Mobility Tactical Truck (HEMMT)- Initial Production Vehicles," Technical Report GL-85-4, U.S. Army Waterways Experiment Station, Vicksburg, MS.

INTENTIONALLY LEFT BLANK.

CRITERION-FREE CURVE FITTING

Dr. Mel Brown
US Army Research Office
Attn: AMXRO-RT
PO Box 12211
Research Triangle Park, NC 27709-2211

ABSTRACT

A non-standard method, called Criterion-Free Curve Fitting (CFCF) is developed for fitting mathematical functions to statistical data consisting of x_i,y_i pairs (generally, n-tuples). CFCF is more flexible than standard linear methods that minimize sums of squares of residuals. Also, CFCF is robust to data outliers, to the presence of errors that are non-normal or heteroscedastic, and to whether independent data (as well as dependent data) variables may also contain measurement errors.

The present paper has been modified in some respects from what was presented at the Conference, as follows: (1) Numerical results are included; (2) Simultaneous solutions are recommended for computing the candidate parameter values; (3) Confidence intervals are discussed and; (4) "Quasi-Median Point Estimation," advocated at the Conference, is discussed but not recommended, as numerical experiments fail to support their effectiveness in fully removing the influence of data outliers.

INTRODUCTION

Traditional curve fitting to statistical data relies on least-squares regression estimation of parameters that select a curve (or surface) from a family of curves (surfaces) corresponding to the possible parameter values. Least squares curve fits are particularly elegant and computationally efficient when the curves are linear. When the data errors are also normal, independent, and homoscedastic, the error analyses are particularly tractable in both a symbolic and computational sense, and the least squares method is provably optimal. The theoretical and computational advantages of the method under idealized data conditions provide a strong incentive to transform variables, where feasible, to linearize the function to be fit. Accordingly, linearization is very common in practice. However, linearization typically comes at the cost of optimality of the fit, and of a reliable and meaningful error analysis. Moreover, linearization is not always feasible.

Even though least squares curve fitting, with or without the help of linearization, is convenient and therefore very common in practice, it remains fragile with respect to the presence of data outliers, and to model assumptions as to normality, independence, and homoscedasticity of errors, and also to whether measurement errors may be present in the independent variables as well as the dependent variables. Moreover, the number of useful linearizing transformations is itself limited, so that least squares curve fits must necessarily remain limited in application.

What is needed is a curve fitting methodology that is flexible (i.e., widely applicable) and robust (where least squares fails to be robust). To meet this need we will introduce an unconventional, heuristic, computationally intensive approach to curve fitting, called Criterion-Free Curve Fitting (CFCF). CFCF, typically, requires repeated numerical solutions of systems of possibly nonlinear equations (K equations if there are K parameters to fit). In principle, this can often be exceedingly complex and impractical, especially if the process is to be automated and repeated many times (as CFCF indeed

Approved for public release; distribution is unlimited.

requires). Fortunately, it often happens in practical applications that additional information is available in the form of initial ballpark estimates, or perhaps constraints, that can help to identify the solution in an acceptably efficient manner. CFCF therefore promises to provide a highly practical approach to curve fitting in many cases where traditional methods fail.

It turns out that CFCF requires that point estimates be generated from random sample distributions. This will lead us to an explore the merits of some candidate point estimators.

THE PROBLEM

We want to fit $y = f(\underline{x};\underline{a})$ from a family of functions characterized by parameters a_m to data points (x_i,y_i), or equivalently, $f(\underline{x},y;\underline{a}) = 0$ from a similarly characterized family of equations.

We want to do this without minimizing

$$\sum (residuals)^2 \text{ or } \sum |residuals| \text{ or } median\{|residual|\}$$

or other such criterion.

We will accomplish the necessary fit via a method which we call Criterion-Free Curve Fitting (CFCF), which is explained in the following section via a combination prototype and example.

CRITERION-FREE CURVE FITTING: EXAMPLE AND PROTOTYPE

We choose for illustration the simplest non-trivial example:

$$y = a + bx$$

which we already know how to fit by any of several traditional techniques.

The CFCF procedure works as follows:

Choose an arbitrary pair of distinct indices, (i,j). (If there were K parameters to fit we would choose K distinct indices. Here, K=2)

If the data were error-free we would have

$$\begin{cases} y_i = a + bx_i \\ y_j = a + bx_j \end{cases}$$

and we could solve for the parameters a,b. We will do this, as if errors were not present. However, as the data are presumed to contain errors, this procedure will instead merely yield candidate estimates a_{ij}, b_{ij} indexed on the points that generated them. (In general we would have K parameters, and each candidate parameter value would therefore need to be characterized by K distinct indices.)

We next repeat the process with another pair (i,j), again obtaining candidate estimators a_{ij}, b_{ij}. We keep repeating the process until either: (1) All distinct pairs (K-tuples) have been exhausted or; (2) the distributions of the candidate values have all stabilized. In the latter case, we must select the (i,j) pairs randomly.

Note that in a more general case we would have a system of K non-linear equations to solve for the K candidate parameter estimators. Assuming that solution techniques are available to solve the particular equations at hand, it may still happen that the errors present in the K x,y data points may be such that

these data points fail to lie on one of the assumed family of parametrized curves. In such a case (by definition) solutions for the candidate parameter values cannot be found. If/when this happens, the recommended procedure is to simply ignore this particular K-tuple of x,y values and to go on another set.

The estimators $\hat{a},\hat{b}$ are then defined as point estimators from the respective distributions of the candidate values $\{a_{ij}\},\{b_{ij}\}$.

It might seem that the means of these distributions would be the most appropriate point estimators. In the illustrative example, however, the set of candidate slopes b_{ij} will include some slopes that happen to be between nearby points, each of which is noisy. Some of those candidate b_{ij} are therefore likely to contain large errors and therefore to behave as outliers that might disproportionately contaminate the sample means. Any genuine outliers that may be present in the original data would further contaminate the means, probably significantly. In contrast, the medians are expected to be robust, both to noisy data and to outliers.

The issue is explored numerically in Tables 1 and 2. Table 1 presents results for the prototype example,

$$y = a + bx + \varepsilon$$

$$\text{with } a = 1, b = 1, \varepsilon = Normal(\mu = 0, \sigma^2 = 1).$$

This is a standard textbook case for which linear least squares regression is provably optimal. Table 1 verifies this, and further shows that CFCF also gives reasonably good results (i.e., close to least squares) when based on the median (rather than the mean) for point estimation, except when the number of data points is small (e.g., 5).

Table 2 explores the same case, modified so that simulated outliers have been added randomly with probability 0.05 (i.e., each data point has this probability of having an outlier added to it). Each outlier is generated as a normally distributed random variable with a mean of 0 and a variance of 100. For this case, linear least squares regression is no longer optimal, but is seriously degraded by the outliers. CFCF based on the mean is also degraded. However, CFCF based on the median continues to give good results, essentially indistinguishable from those for which there are no outliers (as in Table 1).

Table 1

Numerical Results for the Simple Case, $y = a + bx + \varepsilon$
with $a = 1, b = 1, \varepsilon = Normal(\mu = 0, \sigma^2 = 1)$
(100 replications of each solution. Sampling of data is with replacement.)

		RMS Errors					
		$\hat{a}$			$\hat{b}$		
No. of data points	Sample Size, (i,j) pairs	Linear Regression	CRCF (Mean)	CFCF (Median)	Linear Regression	CRCF (Mean)	CFCF (Median)
5	20	1.04	1.35	1.30	.33	.40	.39
10	20	.60	1.24	1.07	.10	.20	.16
	50	.64	.90	.80	.11	.14	.13
	100	.74	.91	.86	.12	.13	.13
25	50	.45	1.13	.69	.028	.071	.043
	300	.36	.53	.49	.026	.039	.030
50	100	.33	1.08	.48	.011	.034	.015
	400	.30	.64	.36	.010	.021	.011

Table 2

Numerical Results for the Simple Case, $y = a + bx + \varepsilon + \psi$ with Outliers

with $a = 1, b = 1, \varepsilon = Normal(\mu = 0, \sigma^2 = 1)$

and Outliers $\psi = \begin{cases} Normal(\mu = 0, \sigma^2 = 100), probability = 0.05 \\ 0, probability = 0.95 \end{cases}$

(100 replications of each solution. Sampling of data is with replacement.)

		RMS Errors					
		$\hat{a}$			$\hat{b}$		
No. of data points	Sample Size, (i,j) pairs	Linear Regression	CFCF (Mean)	CFCF (Median)	Linear Regression	CFCF (Mean)	CFCF (Median)
5	20	2.36	2.75	2.27	.66	.74	.59
10	20	1.90	3.97	1.04	.32	.51	.17
	50	1.40	2.42	.94	.24	.33	.15
	100	1.80	2.41	.86	.30	.38	.15
25	50	1.03	2.80	.68	.08	.18	.04
	300	.96	1.27	.50	.07	.09	.04
50	100	.66	3.22	.50	.02	.10	.02
	400	.66	1.44	.41	.02	.05	.01

POINT ESTIMATION

Given the results shown in Tables 1 and 2, it is tempting to try to devise a point estimator that would be numerically close to the mean when the data is ideal, but that would remain insensitive to outliers and to other departures from ideal data. Such an estimator would therefore share in most of the advantages of both the median and the mean.

One attempt to construct such an estimator is described in the following section. The proposed estimator is called, for lack of a better name, the Quasi-Median. Unfortunately, numerical exploration shows that the Quasi-Median fails to fully remove the influence of data outliers. Results are therefore not included in Tables 1 and 2, and the Quasi-Median is accordingly not recommended in practice.

It is possible that a trimmed median might work better, but this alternative has not been explored.

QUASI-MEDIAN POINT ESTIMATION

We begin by assuming that we have N data points denoted by x_i. Note that these x_i have nothing to do with the (x_i,y_i) data points that we have been considering. They are simply any data sample for which we seek a point estimate. Assume further that the data are sorted so that

$$x_1 \le x_2 \le \cdots \le x_N$$

Define also the median index, m, by

$$m = \frac{(N+1)}{2}$$

If N is even, then there is no x_m. However, this will not cause a problem, because we will not be needing a value for x_m.

Ideally, what we would like is to take the mean of the presumably "good" points in the middle of the distribution, and discount the presumably "bad" points far from the middle. To approximate this, we define the Quasi-Median, Q, as

$$Q = \sum_i w_i x_i$$

where the w_i are weights, large near the median and small far from the median.

There are many ways to define such weights. One way is as follows:

$$w_i = cR^{-\left(\frac{2}{N-1}\right)^2 (m-i)^2}$$

where R is an exogenously chosen "influence ratio" defined as

$$R = \frac{w_m}{w_1} \qquad \left(= \frac{w_m}{w_N}\right) \qquad \text{(e.g., } R = 100\text{)}$$

and c is chosen to make

$$\sum_i w_i = 1$$

With this definition, it is readily shown that

$$\begin{cases} \lim_{R\to 1} Q = mean \\ \lim_{R\to \infty} Q = median \end{cases}$$

As a practical matter, one would want R to be big enough to discount outliers near i=1, i=N, but small enough to allow contributions from data near i=m.

ASSESSMENT OF QUASI-MEDIAN POINT ESTIMATION

Unfortunately, numerical experiments show that the Quasi-Median works about as well as the median, but not better, when the data is satisfies the standard assumptions and no outliers are present. With outliers, the Quasi-Median fails to discount the outliers fully, and as a consequence is inferior to the median. It is possible that a trimmed median might work better, but this was not explored.

Accordingly, the Quasi-Median, though perhaps interesting in its own right and potentially useful in other circumstances, cannot be recommended for use in the CFCF method.

CONFIDENCE INTERVALS FOR THE PARAMETER ESTIMATES

Confidence intervals for the CFCF parameter estimates are exceptionally easy to determine. This is a consequence of the fact that the CFCF procedure generates statistical distributions for each of the candidate parameter estimates. These distributions can be displayed as histograms, and confidence intervals can be read directly from them (with interpolation as may be necessary), for any specified confidence level. Such confidence intervals can be symmetric or asymmetric, as needed.

Consistent with the heuristic nature of CFCF, no explicit theories (or fragile assumptions) are needed to estimate confidence intervals.

REPRESENTATIVE APPLICATIONS OF CFCF

CFCF, though inherently robust, is also more flexible than standard linear least squares curve fitting procedures. To illustrate this flexibility, we will outline how CFCF can be used to fit data to mathematical functions that do not lend themselves to linearization, and for which standard methods therefore fail. Two examples will be presented.

EXAMPLE 1

Our first example is

$$y = A(x+c)^b$$

which is a generalization of the standard linearizable case for which $c = 0$. This more general case does not lend itself to linearization by taking logarithms, because $\ln(x+c)$ does not simplify into anything tractable for use in a least squares minimization. In contrast, CFCF is relatively straightforward.

We start by choosing indices i,j,k (because there are 3 parameters to determine). Then

$$\begin{cases} y_i = A(x_i+c)^b \\ y_j = A(x_j+c)^b \end{cases}$$

Divide, take logarithms, and rearrange to get

$$b = \frac{\ln y_i - \ln y_j}{\ln(x_i+c) - \ln(x_j+c)}$$

Repeat for indices i,k. Comparing, we eliminate b and get

$$\frac{\ln(x_i+c) - \ln(x_j+c)}{\ln(x_i+c) - \ln(x_j+c)} = \frac{\ln y_i - \ln y_j}{\ln y_i - \ln y_k}$$

or, with some rearranging,

$$(x_i+c)^{\ln\left(\frac{y_j}{y_k}\right)} (x_j+c)^{\ln\left(\frac{y_k}{y_i}\right)} (x_k+c)^{\ln\left(\frac{y_i}{y_j}\right)} = 1$$

This can be solved numerically for c. At a minimum, the existence of a solution will require that all three y_i be of the same sign. Numerical experimentation with this expression suggests that the expression

is well behaved and gives a unique solution (when a solution exists) that can readily be found by any of several standard numerical techniques. Given a solution c_{ijk}, one can solve readily for b_{ijk} and then for A_{ijk}.

Repeating as necessary for other i,j,k leads to distributions of the $\{A_{ijk}\},\{b_{ijk}\},\{c_{ijk}\}$, from which point estimates $\hat{A},\hat{b},\hat{c}$ (e.g., the sample medians) and confidence intervals can be obtained numerically.

EXAMPLE 2

Our second example is

$$y = A\sin(\omega x + b)$$

If there were sufficient data, with the x_i equally spaced, we might be able to identify ω as the dominant frequency obtained from a Fourier transform of the y_i data. Even then, it is not clear how we would obtain estimates for A,b. In contrast, the CFCF method does not make any special demands on the data, and provides estimates for all three parameters.

We proceed by choosing a triplet i,j,k of data, whereupon

$$\frac{y_i}{y_j} = \frac{\sin(\omega x_i + b)}{\sin(\omega x_j + b)} = \frac{\sin(\omega x_i)\cos b + \cos(\omega x_i)\sin b}{\sin(\omega x_j)\cos b + \cos(\omega x_j)\sin b}$$

Rearranging, we obtain

$$\tan(b) = \frac{y_i \sin(\omega x_j) - y_j \sin(\omega x_i)}{y_j \cos(\omega x_i) - y_i \cos(\omega x_j)}$$

Repeating for indices i,k and comparing to eliminate b, we have

$$\frac{y_i \sin(\omega x_j) - y_j \sin(\omega x_i)}{y_j \cos(\omega x_i) - y_i \cos(\omega x_j)} = \frac{y_i \sin(\omega x_k) - y_k \sin(\omega x_i)}{y_k \cos(\omega x_i) - y_i \cos(\omega x_k)}$$

With some rearranging and simplification we can write this as

$$y_i \sin[\omega(x_j - x_k)] + y_k \sin[\omega(x_i - x_j)] + y_j \sin[\omega(x_k - x_i)] = 0$$

which can be solved numerically for ω_{ijk}, provided the particular ijk data admits of a solution. Note that $\omega = 0$ is always a solution, though (usually) spurious. Numerical experimentation with this expression shows that we seek (via any of several standard numerical techniques) the smallest solution for ω in the half-open interval $0 < \omega \le 2\pi$ (if 2π is the smallest solution, replace it by 0). Given the solution, we can readily backtrack to solve for $\tan(b)$, and thus b_{ijk}. Further backtracking gives us A_{ijk}.

Repeating as necessary for other i,j,k leads to distributions of the $\{A_{ijk}\},\{b_{ijk}\},\{\omega_{ijk}\}$, from which point estimates $\hat{A},\hat{b},\hat{\omega}$ (e.g., the sample medians) and confidence intervals can be obtained numerically.

CONCLUSIONS

A non-standard, computationally intensive heuristic method, called Criterion-Free Curve Fitting (CFCF), has been developed for fitting mathematical functions to statistical data consisting of x_i,y_i pairs (generally, n-tuples). CFCF is more flexible than standard linear methods that minimize sums of squares of residuals. Also, CFCF is robust to data outliers, to the presence of errors that are non-normal or heteroscedastic, and to whether independent data (as well as dependent data) variables may also contain

measurement errors. CFCF readily lends itself to numerical estimates for confidence intervals of all estimated parameters.

The CFCF method requires point estimates to be made from sample data distributions. Over conditions that may include data outliers or other non-ideal data features, the median appears to be preferable to the mean as an estimator. An attempt was made to devise a point estimator, called the Quasi-Median, that might preserve the best features of both mean and median. Unfortunately, the Quasi-Median fails to live up to its intended purpose of fully removing the influence of data outliers when these are present. The Quasi-Median is therefore not recommended. It is possible that a trimmed mean might work well, but this possibility was not investigated.

As part of the computationally intensive process, it is in general necessary to solve a system of K (possibly non-linear) equations for candidate values for the K parameters. This is difficult to do as part of a general-purpose software package. It is likely, therefore, that in the foreseeable future it will be necessary to write special-purpose software to solve these equations, or to adapt available "general purpose" equation solvers to repetitive solutions. In some cases, prior knowledge of "ballpark values" of the parameters may be needed, either to ensure that the solution algorithm converges to a solution or to ensure that it identifies the correct solution.

CFCF is therefore unlikely in the foreseeable future to be included as part of a statistical package that could be friendly even to the mathematically unsophisticated orcasual user. In contrast, CFCF is more likely to remain foreseeably in the toolbox of the professional statistician. This is perhaps ironic, given the conceptual simplicity of the method and its freedom from sophisticated assumptions and theoretical justification.

INTENTIONALLY LEFT BLANK.

INDIVIDUAL BIOEQUIVALENCE: THE BIOEQUIVALENCE SENSITIVITY RATIO COMPARES CRITERION OF BIOAVAILABILITY MERIT

Marshall N. Brunden and Thomas J. Vidmar
Biomathematics Group, Pharmacia and Upjohn Incorporated
Kalamazoo, MI 49001, USA

ABSTRACT

The sensitivity ratio, used in physical science to compare competing methods of measurement, is applied to the bioequivalence problem in pharmaceutical science. Simply stated the sensitivity ratio is the ratio of the true process variability expressed as a function of the measured process variability via a delta method argument. As applied to the bioavailability problem, the AUC serves as a measurement of an individuals' true bioavailability for a formulation. The standard deviation of this true bioavailability provides a criterion for the quality, or merit, of this formulation with regard to bioavailability. The ratio of test to reference formulation merit results in a bioequivalence sensitivity ratio (BSR). This quantity may be used to estimate individual bioequivalence as the BSR parameters are in terms of the marker measurements. The test formulation is inferior to the reference when this ratio is appreciably less than unity. Application is made to a data set from the literature.

INTRODUCTION

Manufacturers who wish to market new formulations of approved drugs must establish that the new drug is bioequivalent to the old. It is well recognized that while population bioequivalence requires that the average bioavailability of two compounds be sufficiently close (Sheiner [1]), more is needed. One would like to conclude that the two formulations have similar distributions as well.

We consider the case of a reference (R) and test (T) formulation from which we may obtain subjects' measurements of bioavailability parameters (AUC, etc.). There is an obvious distinction among the property B (true bioavailability) to be measured and the actual bioavailability marker measurement Y made for that purpose. In actuality there is more than one type of marker measurement for assessing bioequivalence.

The problem is of comparing the two compounds' relative merits and the more basic question of whether the test formulation is bioequivalent to the reference formulation. The choice between the two formulations is not only a technical question but is also dictated by medical and economic factors.

Initial forays [2,3] into assessing the bioequivalence among two formulations concentrated singly upon acceptable differences between the two populations' (μ_T *and* μ_R) mean bioavailability markers. With use of the error variance from an (*ANOVA*) and the difference in sample means of the reference and test formulations, a $(1 - \alpha)\times 100\%$ confidence interval is formed for the difference in population means, μ_T *and* μ_R resulting in $k_1 < (\mu_T - \mu_R) < k_2$. Clinically, it will have been decided that the two formulations can be considered bioequivalent if $K_1 < (\mu_T - \mu_R) < K_2$. The decision rule is to accept bioequivalence if $k_1 > K_1$ and $k_2 < K_2$. Hauck and Anderson [4] reformulated the problem by making nonequivalence the null hypothesis and bioequivalence the alternative hypothesis using the population means μ_T *and* μ_R . Their statistic results in a noncentral t - distribution. Replacing the sample estimate s, for the unknown population standard deviation σ, allows treatment of the noncentrality parameter as approximately a known constant. Consequently the problem can be reformulated by using the central t - distribution.

The recent concept of individual bioequivalence is discussed by Anderson and Hauck [5] . These authors define bioequivalence in an individual j as $1 - K_I \le Y_{Tj} / Y_{Rj} \le 1 + K_I$, where Y_{ij} is the measured bioavailability marker of formulation i (i = R,T) in the j^{th} subject and K_I is the equivalence criteria. They let P_E be the population

Approved for public release: distribution is unlimited.

proportion of subjects where the two formulations are individually bioequivalent and MINP the minimum acceptable proportion that must be bioequivalent in order to call the formulations bioequivalent. They then test H_0: $P_E \leq MINP$ *versus* H_A: $P_E > MINP$.

A very recent paper by Schall and Luus [6] addresses population and individual bioequivalence through difference among the population means μ_T *and* μ_R . Their general idea is to use a comparison of the reference formulation to itself (repeatability) as a basis for the comparison of the test with the reference formulation. They then evaluate this new criteria by use of bootstrap confidence intervals.

More recently Esinhart and Chinchilli [7] provided an extension of the Anderson and Hauck [5] method by applying tolerance intervals.

We propose a criterion of bioavailability merit for assessing the individual bioequivalence. This criteria does not involve the transformation of the marker measurements into dichotomies based upon a definition of bioequivalence and the subsequent establishment of the minimum acceptable proportion as in Anderson and Hauck[5]. Neither does this method require the that the references formulation be compared to itself as done by Schall and Luus [6] .

The definitions of criterion of bioavailability merit and bioequivalence sensitivity ratio are not new and are paraphrases of John Mandel's [8] definitions of a criterion of technical merit and relativity sensitivity. He considers two measurement processes $\mathbf{M_1}$ and $\mathbf{M_2}$ to determine the same property **Q** and develops a sensitivity ratio comparing one process to the other.

We consider two formulations of a drug used in a bioavailability measurement process for $\mathbf{Y_R}$(reference) and $\mathbf{Y_T}$(test). Both $\mathbf{Y_R}$ and $\mathbf{Y_T}$ are functions of a subjects bioavailability potential **B**, for the drug. Since $\mathbf{Y_R}$ and $\mathbf{Y_T}$ are functions of the same **B**, they also are functionally related to each other. We will show how the sensitivity ratio can be used to compare two formulations' individual bioavailability when there is acceptable population bioavailability.

STANDARD DEVIATION OF PREDICTED BIOAVAILABILITY POTENTIAL

Let $\mathbf{Y_i}$, **i=R,T** represent the measure marker value of bioavailability and **B**, an individual characteristic that we define as the subject bioavailability potential for a drug. At this point our presentation is conceptual but it is recognized that **B** may be represented by a function of those constants related to the modeling of drug concentration in the individual. A relationship must exist between **Y** and **B**:

$$Y_i = f_i(B) \qquad (1)$$

f_i is considered to be a differentiable function of **B** such that $f_i'(B) \neq 0$ for every **B** on a given interval of the real line.

If we had the calibration curve of marker Y_i in terms of bioavailability potential B , we could use it to estimate a value of B_i , given by $\hat{B}_i$ for any measurement value Y_i .

In general

$$\hat{B}_i = f_i^{-1}(Y_i) = g_i(Y_i)$$

Where g_i is defined on the range of f_i .

Let ε_i represent the error of measurement of marker $\mathbf{Y_i}$, then applying the law of propagation of error [8] we have

$$\sigma_{\hat{B}_i} \approx \left| \frac{dg_i(Y_i)}{dY_i} \right| \sigma_{\epsilon_i} \qquad (2)$$

The above expression is a valid approximation if the error ε_i in $\mathbf{Y_i}$ is small corresponding to $\mathbf{Y_i}$ (ie. $\varepsilon_i << Y_i$). If $Y_i = f_i(B)$ is a linear function in **B** the expression given by (2) is exact.

Since $g_i = f_i^{-1}$ is the inverse function of f_i,

$$\frac{dg_i(Y_i)}{dY_i} = \frac{1}{\frac{df_i(B)}{dB}}$$

Then inserting this result into (2) we have:

$$\sigma_{\hat{B}_i} \approx \frac{\sigma_{\epsilon_i}}{\left|\frac{df_i(B)}{dB}\right|} \qquad (3)$$

Surprisingly we observe that this equation allows us to convert the standard deviation of the measured marker Y_i, σ_ϵ, into a standard deviation of the individual "estimated" bioavailability potential $\hat{B}_i$, although the actual estimate is not necessary. What is required is a knowledge of the tangent to the unknown calibration curve. We will see in comparing two formulations' standard deviations of individual "estimated" bioavailability potential that this will not cause us a problem.

COMPARISONS OF ESTIMATED INDIVIDUAL BIOAVAILABILITY POTENTIAL

Consider the two measurement processes Y_T and Y_R associated with the test and reference formulations. To "estimate" the subject bioavailability potential B, we will use the two calibration functions given in equation (1).

Define the formulation with the greater bioavailability merit as the one that has the smaller $\sigma_{\hat{B}_i}$. Consider next the ratio of the two standard deviations of "estimated" individual bioavailability merit given by (3):

$$\left(\frac{\sigma_{\hat{B}_R}}{\sigma_{\hat{B}_T}}\right) \approx \left(\frac{\sigma_{\epsilon_R}}{\sigma_{\epsilon_T}}\right)\left|\frac{\frac{df_T(B)}{dB}}{\frac{df_R(B)}{dB}}\right| \qquad (4)$$

Since the marker responses of the reference and test formulations are functions of the same subjects' bioavailability potential **B**, both Y_T and Y_R must be related to each other.

If Y_T is plotted versus Y_R then the derivative $\frac{dY_T}{dY_R}$ may be written as $\frac{\frac{dY_T}{dB}}{\frac{dY_R}{dB}}$. Differentiating equation (1) and substituting we have:

$$\frac{\frac{dY_T}{dB}}{\frac{dY_R}{dB}} = \frac{\frac{df_T(B)}{dB}}{\frac{df_R(B)}{dB}}$$

and equation (4) becomes:

$$BSR\left(\frac{Y_T}{Y_R}\right) = \left(\frac{\sigma_{\hat{B}_R}}{\sigma_{\hat{B}_T}}\right) = \left(\frac{\sigma_{\epsilon_R}}{\sigma_{\epsilon_T}}\right)\left|\frac{dY_T}{dY_R}\right| \qquad (5)$$

This is the bioequivalence sensitivity ratio (BSR) of the test formulation with respect to the reference formulation. Remarkably, this ratio of the standard deviations of the individual "estimated" bioavailability potential estimates, $\hat{B}_i$ associated with the two formulations can be expressed in terms of parameters related to the two marker measurements Y_T and Y_R, without having the calibration curves of Y_T and Y_R in terms of **B**.

The following is taken from Mandel[8]. Consider Figure 1 where the AUC's for subjects A & B for both the reference and test formulations are displayed. The relationship between the reference and test formulation is represented by the curve. Based on this figure, we would state that the reference formulation is better able to

differentiate between subjects' A & B since the change in the AUC for the reference formulation, call it ΔY_R is larger than the change in the test formulation, ΔY_T . However, the arguement does not consider the error involved in measuring the AUC for the reference or test compound. Therefore, a better comparison would be $\left(\frac{\Delta Y_T}{\sigma_{\varepsilon_T}}\right)\Big/\left(\frac{\Delta Y_R}{\sigma_{\varepsilon_R}}\right) = \left|\frac{\Delta Y_T}{\Delta Y_R}\right|\left(\frac{\sigma_{\varepsilon_R}}{\sigma_{\varepsilon_T}}\right)$, the absolute value to take into account a decreasing curve. Note that for points A & B that are fairly close to one another, what we have described is simply equation (5).

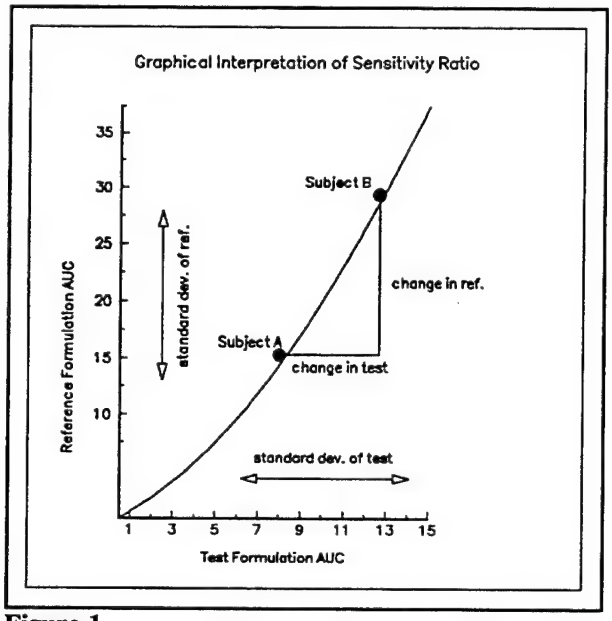


Figure 1

In equation (5) as $\sigma_{\hat{\beta}_R}/\sigma_{\hat{\beta}_T}$ increases the bioavailability merit of the reference formulation decreases with respect to that of the test formulation. If this ratio is appreciably less than unity, the test formulation is "technically" inferior to the reference formulation. Then of the two formulations, the reference formulation has a greater ability to detect a real difference in subject bioavailability potential (individual bioequivalence).

Unless the relationship between $\mathbf{Y_T}$ and $\mathbf{Y_R}$ is linear (*ie.* $\frac{df_T(B)}{df_R(B)} = c$, *a constant*), the slope will vary along the curve of $\mathbf{Y_T}$ versus $\mathbf{Y_R}$. The standard deviations of σ_{ε_R} and σ_{ε_T} may also not be constant throughout the range of variation of $\mathbf{Y_T}$ and $\mathbf{Y_R}$. By plotting the bioequivalence sensitivity ratio versus $\mathbf{Y_T}$ one obtains a complete picture of the individual bioavailability merit of the test formulation relative to the reference formulation (ie. the bioequivalence sensitivity curve of the test formulation relative to the reference formulation).

It should also be noted that the value of the individual bioequivalence sensitivity ratio is invariant [8] with respect to any transformation of scale (eg. if we took logarithms of the $\mathbf{Y_R}$ measurements). The proof is simple and involves using the transformation $\mathbf{Y_R}^* = T(\mathbf{Y_R})$. Next $d\mathbf{Y_R}^*/d\mathbf{Y_T}$ and σ^*_{YR} are substituted into $BSR(\mathbf{Y_R}^*/\mathbf{Y_T})$ which is shown to reduce to $BSR(\mathbf{Y_R}/\mathbf{Y_T})$.

STRATEGIES FOR TESTING OF INDIVIDUAL BIOEQUIVALENCE

USE OF THE REFERENCE FORMULATION COMPARED TO ITSELF

In summary, we have demonstrated that conditional upon an acceptable population bioequivalence, the individual bioequivalence, as measured through a bioavailability marker, may be assessed through the use of a criterion of bioavailability merit and a bioequivalence sensitivity ratio. The result is the ratio of the standard deviations of

individual "estimated" bioavailability potential of the test formulation to the reference formulation for the drug in question.

One could consider using the bioequivalence sensitivity measure to first compare the reference formulation to itself (second trial) as the basis for evaluating the bioequivalence sensitivity measure comparison of the test formulation to the reference formulation. This strategy was recommended by Schall and Luus [4] when using the differences between two bioavailabilities. However, it is easily shown that this strategy offers no advantage when using the bioequivalence sensitivity measure.

Let $BSR(Y_{R'}/Y_R)$ denote the bioequivalence sensitivity of the reference formulation with itself and $BSR(Y_T/Y_R)$ the bioequivalence sensitivity of the test formulation relative to the reference formulation.

Then if $BSR\left(\frac{Y_T}{Y_R}\right)/BSR\left(\frac{Y_{R'}}{Y_R}\right) \geq K$, $0<K\leq 1$, we say the test formulation has acceptable bioequivalence.

The particular value for K, which defines acceptable bioequivalence would be a regulatory concern.

It is easy to show that the BSR of the reference with itself has no dividend since

$BSR\left(\frac{Y_T}{Y_R}\right)/BSR\left(\frac{Y_{R'}}{Y_R}\right) = BSR\left(\frac{Y_T}{Y_{R'}}\right)$, and only the latter realization is necessary to test for bioequivalence.

DEFINITION OF ACCEPTABLE INDIVIDUAL BIOEQUIVALENCE

If $BSR\left(\frac{Y_T}{Y_R}\right) \geq K$, $0<K\leq 1$, then the test formulation has acceptable bioequivalence. It is easily shown that this is equivalent to $\frac{\sigma_{\hat{B}_T}}{\sigma_{\hat{B}_R}} \leq \frac{1}{K}$, $1\leq\frac{1}{K}\leq\infty$. That is, when the standard deviation of the test formulation bioavailability is acceptable in relation to the standard deviation of the reference formulation bioavailability.

Say, that the value of K is chosen such that the bioequivalence sensitivity of test with respect to the reference is .80 of the bioequivalence sensitivity of the reference with respect to itself. K=.8 translates into a standard deviation of the test formulation "estimated" bioavailability potential which may be up to $1/K = 1.25$ larger than the standard deviation of the reference formulation "estimated" bioavailability potential and yet be considered bioequivalent.

MODEL AND NOTATION

We propose the following model for observations from a two period crossover design:

$$y_{ijk} = \mu_i + \pi_k + S_{ij} + \eta_{ijk}$$

$$\textit{where } i = R,T \text{ , } j = 1,\ldots,n \textit{ and } k = 1,2 \text{ .}$$

Where y_{ijk} is the marker measurement (possibly log transformed) taken in the k^{th} period on subject j receiving the i^{th} formulation. We let $E(y_{ijk}|j) = \mu_{ij} + \pi_k = \mu_i + S_{ij} + \pi_k$, *with* $Var\ (y_{ijk}|j) = E(\eta^2_{ijk}) = \sigma^2_{w_i}$. Further, $E_j(\mu_{ij}) = E_j(\mu_i + S_{ij}) = \mu_i$, $Var_j(\mu_{ij}) = E_j(S^2_{ij}) = \sigma^2_{b_i}$ *and* $Cov_j(\mu_{Rj},\mu_{Tj}) = \rho\sigma_{b_R}\sigma_{b_T}$. This model is similar to that proposed by Anderson and Hauck [5] for individual bioequivalence aside from the inclusion of the period effect.

Then, the marker responses for R and T may be written:

$$y_{Rjk} = \mu_R + \pi_k + S_{R_j} + \eta_{Rjk}$$

$$y_{Tjk} = \mu_T + \pi_k + S_{T_j} + \eta_{Tjk}$$

The population expectations are given:

$$E(y_{Rjk}) = \mu_R + \pi_k = \mu_{Rk}$$

$$E(y_{Tjk}) = \mu_T + \pi_k = \mu_{Tk}$$

The population variances are given by:

$$Var(y_{Rjk}) = E(S_{Rj} + \eta_{Rjk})^2 = \sigma_{b_R}^2 + \sigma_{w_R}^2 = \sigma_{\epsilon_R}^2$$

$$Var(y_{Tjk}) = E(S_{Tj} + \eta_{Tjk})^2 = \sigma_{b_T}^2 + \sigma_{w_T}^2 = \sigma_{\epsilon_T}^2$$

Let the subject effects conditional on subject j within each formulation i, be a "structural" component. That is, $S_{ij} = h_i(B_j)$, where B_j is a constant which represents subject j's bioavailability potential for the drug (at a particular dose). The function h_i imparts the marker effect due to formulation i as a function of this bioavailability potential for the drug. To be consistent with our overall model the expectation of $h_i(B_j)$ over the population J of subjects is $E_j(S_{ij}) = E_j[h_i(B_j)] = 0$, the variance of $h_i(B_j)$ over J is $E_j(S_{ij}^2) = E_j[(h_i(B_j))^2] = \sigma_b^2$. , and the covariance of reference and test given by $E_j(S_{Rj}, S_{Tj}) = E_j[(h_R(B_j)),(h_T(B_j))] = \rho\sigma_{b_R}\sigma_{b_T}$.

Now, the marker responses for R and T given subject j in period k may be written:

$$y_{Rjk}|j = \mu_{Rk} + h_R(B_j) + \eta_{Rjk|j}$$

$$y_{Tjk}|j = \mu_{Tk} + h_T(B_j) + \eta_{Tjk|j}$$

The individual expectations are given:

$$Y_{R_{jk}} = E(y_{Rjk}|j) = \mu_{Rk} + h_R(B_j)$$

$$Y_{T_{jk}} = E(y_{Tjk}|j) = \mu_{Tk} + h_T(B_j) \qquad (6)$$

The variances conditional on j are given by:

$$Var(Y_{R_{jk}}) = E(\eta_{Rjk|j})^2 = \sigma_{w_R}^2$$

$$Var(Y_{T_{jk}}) = E(\eta_{Tjk|j})^2 = \sigma_{w_T}^2$$

From (6) we see that the mean bioavailability response from both the reference and test formulations are functions of both the population bioavailability and subject bioavailability potential. This is the functional relationship of Y_T to Y_R among subjects that was alluded to within the introduction. If the population levels of bioavailability μ_R *and* μ_T differ by an acceptable quantity the remaining issue is the individual bioequivalence which is specifically addressed in the next two sections.

EXPERIMENTAL DESIGN AND TESTS

SIMPLE MODEL FOR THE CALIBRATION FUNCTION $f_i(B_j)$ *and* $BSR(\frac{Y_T}{Y_R})$

The subject marker measurement bioavailability potential calibration functions for the reference and test given in equation (6) are compatible with compartmental models used for the pharmacokinetic modeling of drug distribution with first-order output. Many of these models result in a similar expression for the total area under the plasma-level time curve [9]:

$$AUC_{ij} = \frac{Q_i}{C_{ij}}$$

Q_i is the quantity of drug available for formulation i and C_{ij} is the drug clearance for formulation i within subject j. In crossover studies [9] C_{ij} is often considered a constant among formulations (ie. $C_{ij} = C_j$).

Considering the $\log(AUC)$ we then have:

$$Y_{ij} = \log(AUC_{ij}) = \log(Q_i) - \log(C_j).$$

Then Y_{Tj} is a linear function of Y_{Rj} having an intercept ($\log(Q_T) - \log(Q_R)$) and a slope of unity. To allow for testing of a formulation effect on clearance we write:

$$Y_{ij} = \log(Q_i) - \gamma_i \times \log(C_j) = \mu_i + \gamma_i \times (B_j)$$

Here the $\log(C_j)$ may be considered to be the subjects bioavailability potential B_j , for the drug.
The bioequivalence sensitivity ratio (5) is then:

$$BSR\left(\frac{Y_T}{Y_R}\right) = \left|\frac{dY_T}{dY_R}\right|\left(\frac{\sigma_{\epsilon_R}}{\sigma_{\epsilon_T}}\right) = \left|\frac{\gamma_T}{\gamma_R}\right|\left(\frac{\sigma_{\epsilon_R}}{\sigma_{\epsilon_T}}\right)$$

If $\gamma_i = 1$ for i = R,T the ratio of the two formulations standard errors of estimated bioavailability merit is:

$$BSR\left(\frac{Y_T}{Y_R}\right) = \left|\frac{dY_T}{dY_R}\right|\left(\frac{\sigma_{\epsilon_R}}{\sigma_{\epsilon_T}}\right) = \left(\frac{\sigma_{\epsilon_R}}{\sigma_{\epsilon_T}}\right)$$

On the other hand, in bioavailability studies it is often assumed that $\sigma_{\epsilon_R}^2 = \sigma_{\epsilon_T}^2 = \sigma_\epsilon^2$
This results in:

$$BSR\left(\frac{Y_T}{Y_R}\right) = \left|\frac{dY_T}{dY_R}\right| = \left|\frac{\gamma_T}{\gamma_R}\right| \qquad (7)$$

The remainder of this paper will concentrate on estimators for $\left|\frac{\gamma_T}{\gamma_R}\right|$ and their application.

EXPERIMENTAL DESIGN CONSIDERATIONS

Hopefully, transformations of Y_T and Y_R may be found which provide both for homoscedasticity of σ_{ϵ_T} and σ_{ϵ_R} and a linear relationship of Y_T with Y_R. We will reject (individual) bioequivalence if $BSR\left(\frac{Y_T}{Y_R}\right)$ is not sufficiently high (at least K). Then our null hypothesis is H_0: $BSR\left(\frac{Y_T}{Y_R}\right) \leq K$ versus the alternative hypothesis H_a: $BSR\left(\frac{Y_T}{Y_R}\right) > K$, or a lower $(1 - \alpha)\times 100\%$ confidence limit of $BSR\left(\frac{Y_T}{Y_R}\right)$. In section 4 we saw that it was not necessary to compare the reference with itself as a basis for evaluating the bioequivalence sensitivity of the test formulation with respect to the reference formulation. This would imply that only one replication of marker measurements is necessary.

We propose a 2-period crossover design using $n = n_1 + n_2$ subjects (Jones and Kenward [10]):
Is it usually argued that in bioavailability trials [2] no carry over effect exists. The
full model fixed effects are:

Sequence Group	No. Subjects	Period 1	Period 2
1 (Reference, Test)	n_1	$\mu_R + \pi_1$	$\mu_T + \pi_2$
2 (Test, Reference)	n_2	$\mu_T + \pi_1$	$\mu_R + \pi_2$

The appropriate analysis format with expected mean squares is:

Source	Degrees of Freedom	Expected Mean Square
Total	2n - 1	
Between Subjects	$n - 1$	$\sigma_W^2 + 2\sigma_B^2$
Periods	1	$\sigma_W^2 + n\cdot(\pi_1 - \pi_2)^2/2$
Formulations	1	$\sigma_W^2 + n\cdot(\mu_T - \mu_R)^2/2$
Within Subjects	$n - 2$	σ_W^2

POPULATION BIOEQUIVALENCE

Using the error variance from the ANOVA (s_W^2) and the difference in the sample means for the reference and test formulations ($\hat{\mu}_T - \hat{\mu}_R$) we may form a ($1 - \alpha$) confidence interval for the difference in population means $\mu_T - \mu_R$, where typically α=.05,:

$$(\hat{\mu}_T - \hat{\mu}_R) \sim NID\left((\mu_T - \mu_R), \frac{2\sigma_W^2}{n}\right)$$

The (1 - α) x 100% confidence interval is given by:

$$(\hat{\mu}_T - \hat{\mu}_R) \pm t_{n-2}(1 - \frac{\alpha}{2})\sqrt{\frac{2s_W^2}{n}}$$

where t_{n-2} (1 - α/2) is student's t-deviate with (n - 2) degrees of freedom and α is usually .05. If these limits are within an acceptable range K_1, K_2 we have bioequivalence. Where the response is the logarithm of the measurement marker, the antilogarithm of both ($\hat{\mu}_T - \hat{\mu}_R$) and the confidence limits is taken. The resulting interval will appear as an interval for μ_T'/μ_R', where μ_i = exp (μ_i'). .

INDIVIDUAL BIOEQUIVLAENCE

From equation (7) we see that we must estimate $\frac{\gamma_T}{\gamma_R}$ and its standard error. Consider next, a set of pairs $[\Delta_{Tjk}|j, \Delta_{Rjk}|j] = [(Y_{Tjk}|j - \mu_{Tk}), (Y_{Rjk}|j - \mu_{Rk})]$. Since $Y_{ijk}|j = \mu_{ik} + \gamma_i B_j$ and μ_{ik} is the mean of formulation i within period k we have $[\Delta_{Tjk}|j, \Delta_{Rjk}|j] = [\gamma_T\beta_j, \gamma_R\beta_j]$.

There exist corresponding measurements

$$y_{ijk}|j = \mu_{ik} + \gamma_i(B_j) + \eta_{ijk|j} \quad \textit{and}$$

$$\bar{y}_{i.k} = \mu_{ik} + \frac{1}{n}\sum_{j=1}^{n} S_{ij} + \frac{1}{n}\sum_{j=1}^{n}\eta_{ijk} .$$

Let $\delta_{ijk}|j = (y_{ijk}|j - \bar{y}_{ijk}) = \gamma_i\beta_j + \left[\eta_{ijk}|j - \frac{1}{n}\sum_{j=1}^{n} S_{Tj} - \frac{1}{n}\sum_{j=1}^{n}\eta_{Ijk}\right] = \gamma_i\beta_j + E_{ijk|j}$.

Then we may write our model:

$$\gamma_T \beta_j = B_{T/R} \gamma_R \beta_j$$

$$\textit{where we wish to estimate } B_{T/R} = \frac{\gamma_T}{\gamma_R}.$$

$$(\delta_{Tjk}|j, \delta_{Rjk}|j) = (\gamma_T \beta_j, \gamma_R \beta_j) + (E_{Tjk|j}, E_{Rjk|j}) \textit{ where } (\delta_{Tjk}|j, \delta_{Rjk}|j) \textit{ is observed,}$$

$$\gamma_T \beta_j \textit{ is the true value of the dependent variable, } \gamma_R \beta_j \textit{ is the true value}$$

$$\textit{of the independent variable and } (E_{Tjk|j}, E_{Rjk|j}) \textit{ is the the vector of measurement errors.}$$

The above is the classical measurement errors model.

MAXIMUM LIKELIHOOD (ML) ESTIMATION

Kendall and Stuart [11] address this problem within their Functional and Structural Relationship chapter. If we assume that $E(E_{Tjk|j}^2) = \lambda E(E_{Rjk|j}^2)$, *where* $\lambda = 1$, their maximum likelihood estimator of $B_{T/R}$ under normality assumptions is given by

$$\hat{\beta}_{T/R} = \frac{\left(s_{\delta T}^2 - s_{\delta R}^2\right) + \left[\left(s_{\delta T}^2 - s_{\delta R}^2\right)^2 + 4 s_{\delta R \delta T}^2\right]^{\frac{1}{2}}}{2 s_{\delta R \delta T}} \qquad (8)$$

where $s_{\delta_R}^2 = \frac{\sum_{j=1}^{n} \left(\delta_{Rj} - \bar{\delta}_T\right)^2}{n}$, $s_{\delta_T}^2 = \frac{\sum_{j=1}^{n} \left(\delta_{Tj} - \bar{\delta}_R\right)^2}{n}$, *and* $s_{\delta_R \delta_T}^2 = \frac{\sum_{j=1}^{n} (\delta_{Tj} - \delta_T)(\delta_{Rj} - \delta_R)}{n}$. .

Kendall and Stuart [11] also demonstrate the consistency of $\hat{\beta}_{T/R}$.

For confidence interval estimation about $\beta_{T/R}$ define $\beta_{T/R} = \tan(\theta)$ and $\hat{\beta}_{T/R} = \tan(\hat{\theta})$. Then the lower $(1 - \alpha)\times 100\%$ confidence limit about θ is:

$$\theta_l = \hat{\theta} - .5 \sin^{-1}\left[2t\left[\frac{s_{\delta R}^2 s_{\delta T}^2 - s_{\delta R \delta T}^2}{(n-2)\left[\left(s_{\delta R}^2 - s_{\delta T}^2\right)^2 + 4 s_{\delta R \delta T}^2\right]}\right]^{\frac{1}{2}}\right] \qquad (9)$$

where t is the appropriate "students" deviate for (n - 2) degrees of freedom for the confidence coefficient being used. Then for $\beta_{T/R}$, the lower $(1 - \alpha)\times 100\%$ confidence limit is $\tan^{-1}(\theta_l)$.

There are limits on (9) due to the periodicities of the tangent function. The absolute difference between the estimated and actual theta must be less than or equal to .25 radian = 45 degrees for the formulation to hold. In addition, the $\sin^{-1}(\Delta)$ does not exist for $\Delta > 1$ and consequently the confidence interval about theta would not exist for the value of the t-statistic used. This is due to either to small α and/or n. Fuller [12] has also developed an estimator $S_{\hat{\beta}_{T/R}}$ of the standard error of $\hat{\beta}_{T/R}$. The quantity $t = \left(\frac{\hat{\beta}_{T/R} - \beta_{T/R}}{S_{\hat{\beta}_{T/R}}}\right)$ is approximately a N(0,1) random variable and it is suggested, that in small samples, t is approximated by Student's t-distribution with n - 2 degrees of freedom.

BOOTSTRAP ESTIMATION METHODOLOGY

As an alternative to estimating $\beta_{T/R}$ in the formulation $\gamma_T(\beta_j) + E_{Tjk|j} = \beta_{T/R} \cdot \gamma_R(\beta_j) + E_{Rjk|j}$ the bootstrapping methodology of Efron and Tibshirani [13] was adopted. Since both γ_T *and* γ_R are known to be associated with error terms, the method of bootstrapping pairs in a no-intercept regression model was used. Each bootstrap estimate consisted of sampling n pairs with replacement followed by a least squares estimate of $\beta_{T/R}$,

a total of 500 times. In this case, n is the number of pairs in the original data. All calculations were done using the SAS [14] software system.

A robust bootstrap was also implemented and was similar except that each observation was weighted by w_j where $w_j = \min\left(1, \left(2 \cdot (1/n)^{1/2} / |DFFits_j|\right)\right)$ and DFFits$_j$ is the usual diagnostic described by Belsey, Kuh, and Welsch [15]. This particular weight was suggested by Hettmansperger [16]. A weighted regression was then performed on each bootstrap sample.

A lower 95% confidence interval for the bootstrap estimators of $\beta_{T/R}$ was computed by using the bootstrap percentile estimator as described by Efron and Tibshirani [13]. The method uses the same methodology as described in the estimation of $\beta_{T/R}$ with the exception that one thousand samplings were employed to calculate $\beta^*_{T/R}$. The lower 95% confidence bound for $\beta_{T/R}$ is then the 5.0% percentile of the $\beta^*_{T/R}$ distribution. Bootstrap estimation has the added avantage of not having to assume $E(E_{Tjk|j}^2) = E(E_{Rjk|j}^2)$.

EXAMPLE OF THE POPULATION AND INDIVIDUAL BIOEQUIVALENCE CALCULATIONS

VIRGINIA COMMONWEALTH VERAPAMIL DATA

A four sequence, four period, two treatment study was conducted on 23 normal subjects at the Department of Pharmacy and Pharmaceutics at Virginia Commonwealth University to determine if a test formulation of verapamil should be considered bioequivalent to a reference formulation. This data set is analyzed within Esinhart and Chinchilli [7].

We have pointed out it is only necessary to have a two period, two treatment study to use our proposed methodology. For an example of our methodology we will analyze the period 1 & 2 data only.

A univariate, linear model on log(AUC) was constructed with subject effects (df = 22), period effects (df = 1) and formulation effects (df = 1). The test for formulation effects was not significant (p = .5251). Based on least squares the estimated 90% and 95% confidence intervals for average bioavailability (test/reference) are (0.8135 - 1.0981) and (0.7884 - 1.1330). Since the 90% confidence interval lies between 0.80 and 1.20, one would reject bioinequivalence in favor of bioequivalence. This result is similar to that found by Esinhart and Chinchilli [7]. Next we estimate the BSR by ML, bootstrap, and robust bootstrap methods with 95% confidence limits and a bootstrap percentile estimator as outlined within Sections 6.5 and 6.6. This resulted in the estimators shown within Table I.

Table I. Estimators from Virginia Common data

estimation method	$\hat{BSR}$	*Lower 95% C.I.*
M.L.	.861	.608
bootstrap	.689	.547
robust bootstrap	.689	.517

The bootstrap and robust bootstrap estimators are very close, while the ML estimate appears to overestimate the bioequivalence sensitivity ratio. We elected to go with robust bootstrap estimation as it is possible that "outlier" observations may exist with small differences among the test and reference formulations variances. The estimate of BSR indicates that the test formulation is "technically" inferior to the reference formulation. We may say that the standard deviation of the individual bioavailability potential for the test formulation is estimated to be 1/.689=1.45 of the reference formulation.

The lower 95% confidence bound is estimated to be .517. We apply a rule analogous to the rule used for the

90% confidence interval for population bioequivalence (average bioavailability) to the bioequivalence sensitivity ratio. We cannot reject bioinequivalence in favor of bioequivalence as the 95% bound is not greater than or equal to the critical bound of .80. As a consequence we conclude that the test formulation is not individually bioequivalent. This same conclusion was reached by Eisenhart and Chinchilli[7].

See Figure 2 for the data points used and the fitted lines relating the test formulation to the reference formulation. We note that the bootstrap and robust bootstrap lines overlap.

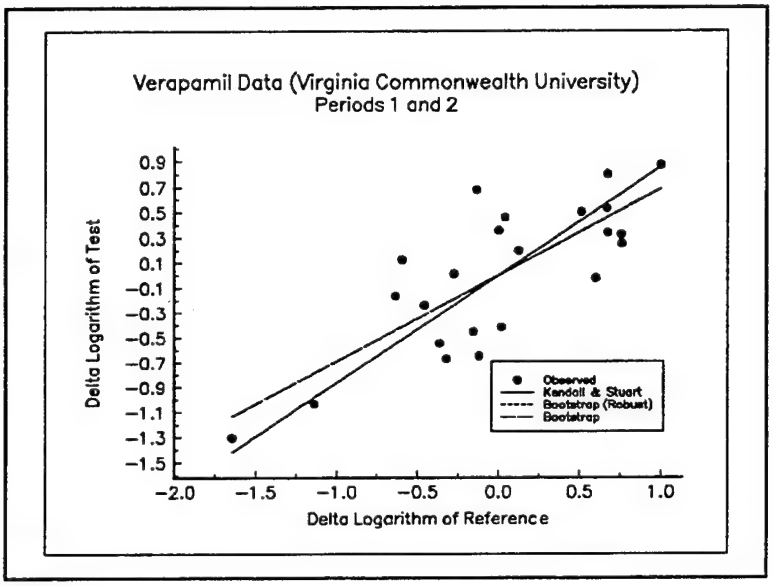


Figure 2

DISCUSSION

Our proposed methodology provides information regarding the similarity of distributions of a test and reference formulations marker values through the use of a criterion of bioavailability merit and the resulting bioequivalence sensitivity ratio. This methodology has the valued property of invariance with respect to any transformation of scale in the marker responses. In addition, only a two period study needs to be conducted as there is no need or advantage in comparing the reference formulation to itself as a basis of comparison of the test with the reference formulation in assessing individual bioequivalence. This leads to both a savings in time and money.

REFERENCES

1. Sheiner, L.B. 'Bioequivalence revisited', *Statistics in Medicine*, **11**, p. 1777, 1992.

2. Westlake, W.J. 'Statistical aspects of comparative bioavailability trials', *Biometrics*, **35**, p. 272, 1981.

3. Westlake, W.J. 'Bioavailability and bioequivalence of pharmaceutical formulations', In Karl Peace (ed.), *Biopharmaceutical Statistics for Drug Development*. Marcel Dekker, New York, p. 329, 1980.

4. Hauck, W.W. and S. Anderson. 'A new statistical procedure for testing equivalence in two-group comparative bioavailability trials', *Journal of Pharmacokinetics and Biopharmaceutics*, **12**, p. 83, 1984.

5. Anderson, S and W.W. Hauck. 'Consideration of individual bioequivalence', *Journal of Pharmacokinetics and Biopharmaceutics*, **18**, p. 259, 1990.

6. Schall, R. and H.G. Luus. 'On population and individual bioequivalence', *Statistics in Medicine*, **12**, p. 1009, 1993.

7. Esinhart, J.D. and V.M. Chinchilli. 'Extension to the use of tolerance intervals for the assessment of individual bioequivalence', *Journal of Biopharmaceutical Statistics*, **4**, p. 39, 1994.

8. Mandel J. *The Statistical Analysis of Experimental Data*, Interscience Publishers, John Wiley & Sons, New York, p. 363, 1964.

9. Portmann, G.A. 'Pharmacokinetics', In James Swarbruck (ed.), *Current Concepts in the Pharmaceutical Sciences: Biopharmaceutics*. Lea and Febiger, Philadelphia, p. 2, 1970.

10. Jones, B. and M.G. Kenward. *Design and Analyses of Crossover Trials*, Chapman and Hall, London, p. 16, 1990.

11. Kendall, M. and A. Stuart. *The Advanced Theory of Statistics - Inference and Relationship*, Volume 2, 4th Edition, Macmillan Publishing Company, New York, p. 389, 1979.

12. Fuller, W.A. *Measurement Error Models*, John Wiley & Sons, New York, p. 30, 1987.

13. Efron, B. and R.J. Tibshirani. *An Introduction to the Bootstrap*, Chapman and Hall, New York, p. 105, 1993.

14. SAS Institute. *SAS/STAT User's Guide, Version 6, Fourth Edition, Volumes 1 & 2*, SAS Institute Inc., Cary, North Carolina, 1990.

15. Belsley, D.A., E. Kuh, R.F. Welsch. *Regression Diagnostics*, John Wiley & Sons, New York, p. 15, 1980.

16. Hettmansperger, T.P. 'Why not try a robust regression ?', Technical Report Number 60, The Pennsylvania State University, 1985.

RELIABILITY ESTIMATES OF COMPLEX STRUCTURES

ASIT P. BASU[1]
UNIVERSITY OF MISSOURI-COLUMBIA

ABSTRACT

In this paper Bayesian methods for estimating reliability of complex structures are considered. The topics considered are reliability estimation of complex systems, especially when there are no failures.

[1]This research has been supported by the U.S. Air Force Office of Scientific Research under grant no. F49620-95-1-0094.

1. INTRODUCTION

Let X, a nonnegative random variable, denote the lifetime of a physical system with cumulative distribution function (cdf) F(x). Then, the mission time reliability, R(x), is the probability that a system will function at mission time x. That is,

$$R_1 = R(x) = P(X > x). \tag{1.1}$$

A second definition of reliability for the stress-strength model is given by

$$R_2 = P(X < Y), \tag{1.2}$$

where X and Y are independent random variables. Here, Y denotes the strength of a component subject to stress X. As an example, let X denote the chamber pressure and Y the burst pressure of a solid propellant rocket engine. The engine is successfully fired if X<Y.

In this paper Bayesian methods for estimating reliability of complex structures are considered. Bayesian estimates for complex systems are considered in Section 2. In Section 3 we consider estimating the probability of a rare event when no failure has occurred.

2. BAYES ESTIMATES

Consider the exponential distribution. Let $\theta = \frac{1}{\lambda}$, $\theta > 0$, $\lambda > 0$. Here, the density function, conditional on expected lifetime $\theta(\theta > 0)$, is given by

$$f(x|\theta) = \frac{1}{\theta} e^{-x/\theta}, \ (x > 0), \tag{2.1}$$

with survival function

$$R(t|\theta) \equiv \bar{F}(t|\theta) = P(X > t|\theta) = e^{-t/\theta}, \ t > 0 \tag{2.2}$$

and the random parameter θ has a given prior distribution $g(\theta)$. Here, $g(\theta)$ is

chosen to reflect prior knowledge about θ. Considerable literature exists discussing inference problems relating to θ and $\bar{F}(t|\theta)$. See, for example, Basu and Tarmast (1987), Balakrishnan and Basu (1995), and Berger (1985).

Let $g(\theta)$ be the conjugate prior, the inverted gamma density

$$g(\theta) = \frac{\alpha^{\upsilon}}{\Gamma(\upsilon)} \theta^{-(\upsilon+1)} e^{-\alpha/\theta}, \ \theta > 0. \tag{2.3}$$

Here, the hyperparameters α and $\upsilon(\alpha, \upsilon > 0)$ are chosen to reflect prior knowledge. Denote this by

$$\theta \sim IG(\alpha, \upsilon). \tag{2.4}$$

Let $\underline{x}=(x_1, x_2,...,x_n)$ be a random sample from $f(x|\theta)$. Then, the posterior distribution of $\theta \sim IG(\alpha + T, \upsilon + n)$, where $T = \sum_{1}^{n} x_i$.

Under squared error loss

$$\hat{\theta} = (\alpha + T)/(\upsilon + n - 1) \tag{2.5}$$
$$Var\{\theta|\underline{x}\} = (\alpha + T)^2 / \left[(\upsilon + n - 1)^2 (\upsilon + n - 2)\right].$$

Posterior bayes estimate of survival function $\bar{F}(t|\theta)$ is

$$\hat{\bar{F}}(t) = \left[1 + t/(\alpha + T)\right]^{-(\upsilon+n)} \tag{2.6}$$

and

$$\mathrm{Var}\{\bar{F}(t|\theta)|x\} = \left(1+\frac{2t}{\alpha+T}\right)^{-(\upsilon+n)} - \left(1+\frac{t}{\alpha+T}\right)^{-2(\upsilon+n)}. \qquad (2.7)$$

For a k-out-of-p system with independent exponential component lifetimes $f(x_i|\theta_i)$, system reliability is

$$R(t|\theta) = \sum_{j=k}^{p}\left\{\sum_{\alpha_j}\left[\prod_{i=1}^{j}\bar{F}\left(t|\theta_{\alpha_j(i)}\right)\prod_{i=j+1}^{p}F\left(t|\theta_{\alpha_j(i)}\right)\right]\right\}.$$

Here, $\sum_{\alpha_j}$ is over all $\binom{p}{j}$ distinct combinations $\alpha_j = (\alpha_j(1), \alpha_j(2), \ldots, \alpha_j(j))$ of the integers {1,2,...,p} taken j at a time such that exactly j of the X_i's are greater than t.

Assume $\theta_i \sim IG(\alpha_i, \upsilon_i)$ i=1,2,...,p. Then, the Bayesian estimates are

$$\hat{R}_S(t) = \prod_{i=1}^{p}\left[1+t/(\alpha_i+T_i)\right]^{-(\upsilon_i+n_i)}$$

for the series (k=p) system. For parallel system

$$\hat{R}_p(t) = 1-\prod_{i=1}^{p}\left\{1-\left[1+\frac{t}{(\alpha_i+T_i)}\right]^{-(\upsilon_i+n_i)}\right\}.$$

3. ESTIMATING PROBABILITY OF FAILURE OF A RARE EVENT

Consider a highly reliable physical system with reliability R=1-P. Here, failure may be a rare event, and the probability P of occurrence of a failure may be quite low. One may want to estimate the probability of failure based on a random sample of size n when a failure has not occurred at all in a random sample of size n. Here, Y, the number of failures, is zero. The maximum likelihood estimate (MLE) of P is zero for all n. This is contrary to our knowledge that the rare event does occur.

The Bayesian estimate of P is a natural one. Basu, Gaylor, and Chen (1996) have considered this problem of estimating the probability of occurrence of tumor for a

rare cancer with zero occurrence in a small sample. Conditional on P, Y follows the binomial distribution with parameters n and P. Denote this by

$$Y|P \sim \text{Bin}(n, P). \tag{3.1}$$

Let the prior distribution of P be beta with parameters a, b. That is, P~B(a, b). The density function of P is given by

$$g(P) = \frac{1}{B(a,b)} P^{a-1}(1-P)^{b-1},\ a, b>0,\ 0<P<1. \tag{3.2}$$

Using Bayes' theorem, the posterior density of P given the data Y=y is given by

$$g(P|y) = \frac{1}{B(a+y, b+n-y)} P^{a+y-1}(1-P)^{b+n-y-1},\ 0<P<1, \tag{3.3}$$

which is B(a+y, b+n-y).

Here, the parameters a and b are chosen to reflect prior knowledge and expert opinion. The informative prior is the conjugate prior for P. If there is no prior opinion, one could use the noninformative prior (Jeffreys' prior) for P with a=b=.5, which is B(.5, .5).

Using the above prior and squared error loss function, the Bayesian estimate of P is given by

$$\hat{P} = \frac{a+y}{a+b+n} \tag{3.4}$$

Note

$$\hat{P} = c\frac{y}{n} + (1-c)\frac{a}{a+b}, \tag{3.5}$$

where $\frac{y}{n}$ is the maximum likelihood estimate, $\frac{a}{a+b}$ is the mean of the prior distribution, and $0<c=\frac{n}{a+b+n}<1$.

If y=0,

$$\hat{P} = \frac{a}{a+b+n}. \tag{3.6}$$

$\tilde{P} \to 0$, the MLE, as $n \to \infty$. From (2.4), the Bayesian estimate of reliability is given by

$$\hat{R} = 1 - \hat{P} = \frac{b+n-y}{a+b+n}. \tag{3.7}$$

Now consider the underlying random variable X, conditional on the parameter θ, to be continuous following the exponential distribution. Here, X denotes the lifetime of a system with conditional density given by (2.1).

Here, reliability at mission time t is (2.2) and probability of failure by time t is

$$P \equiv P(t|\theta) = 1 - \bar{e}^{t/\theta}. \tag{3.8}$$

Let Y denote the number of failures in a random sample of size n and, as before, is Bin(n, P). When there is no failure, using (3.7) and (3.8) with y=0, θ can be estimated from the equation

$$\bar{e}^{t/\hat{\theta}} = \frac{b+n}{a+b+n}.$$

That is,

$$\hat{\theta} = -t/\ln[(b+n)/(a+b+n)]. \tag{3.9}$$

Note, as $n \to \infty$, $\hat{\theta} \to \infty$ indicating that the expected lifetime is quite high.

REFERENCES

Balakrishnan, N and Basu, AP (Eds., 1995). *The Exponential Distribution: Theory, Methods and Applications*, Gordon and Breach, Newark, NJ.

Basu, AP; Gaylor, DW; and Chen, JJ (1996). Estimating the probability of occurrence of tumor for a rare cancer with zero occurrence in a sample, Regulatory Toxicology and Pharmacology (to appear).

Basu, AP and Tarmast, G (1987). Reliability of a complex system from Bayesian viewpoint, *Probability and Bayesian Statistics*, ed. by Viertl, R, 31-38.

Berger, JO (1985). *Statistical Decision Theory and Bayesian Analysis*, Springer-Verlag, NY.

INTENTIONALLY LEFT BLANK.

Using Wearout Information to Reduce Reliability Demonstration Test Time

W. M. Woods
Naval Postgraduate School
Monterey, CA 93943

ABSTRACT

Formulae are developed for the maximum and minimum linear test time required to demonstrate a given lower confidence limit requirement for component reliability whose failure time has a Weibull distribution with known shape parameter, β. When $\beta > 1$, these test times are shown to be significantly smaller than the corresponding required test time under the exponential distribution.

INTRODUCTION

Acronyms

WLT	actual linear test time accumulated on all items tested under the described test plan when failure time of each test item has a Weibull distribution
mWLT, MWLT	[minimum, maximum] value of WLT
ELT	exponential linear test time; value of WLT when $\beta = 1$
LCL, UCL	[lower, upper] confidence limit

Notation

$R(t)$	reliability function
Wei(t; λ, β)	Weibull distribution with $R(t) = e^{-(\lambda t)^{\beta}}$
$\varepsilon(t; \lambda)$	exponential distribution with $R(t) = e^{-\lambda t}$
f	number of failures in sequential time censored test plan
T_i	lifetime of i^{th} component tested in sequential time truncated test plan, $i = 1, 2, \ldots, f$
T_{f+1}	test time accumulated on component number $f + 1$ before the test is terminated
$\chi^2_{\gamma,n}$	100γ percentile point of chi-square distribution with n degrees of freedom
$K(\gamma, f)$	$\chi^2_{\gamma,2(1+f)}/2$

A 100γ% LCL of R_0 for $R(t_0)$ is specified where R_0 and t_0 are given. The time to failure of the device is Wei(t; λ, β) where β is known. The following reliability demonstration test is performed:

> Items are tested sequentially until they fail or the sum of their test times raised to the β power total to a given value T_0^{β} at which time testing is terminated. T_0 is chosen so that if a given number, f, of

Approved for public release; distribution is unlimited.

failures occur in the $\left(0, T_0^{\beta}\right)$ test period, then the computed $100\gamma\%$ LCL for $R(t_0)$ equals R_0. The tests are administered and the number of failures, F, observed. The LCL specification is validated if $F \le f$.

If $\beta \neq 1$, the actual linear test time accumulated on all items tested is random. Minimum and maximum values, mWLT and MWLT, of this random time are derived and their ratio to the linear test time, ELT, ($\beta = 1$), are computed. The results show that approximately one half of the linear test time for $\beta = 1$ is required if $\beta = 1.25$ and $R_0 = .95$ at the 80% level of confidence.

MAIN RESULTS

Suppose time to failure of a device is Wei(t; λ, β), $\beta > 1$, and the time censored reliability demonstration test described in the Introduction section is run to validate a specified $100\gamma\%$ LCL value of R_0 for $R(t_0)$. Then

$$\frac{\text{mWLT}}{\text{ELT}} = \left(\frac{-\ln R_0}{K(\gamma, f)}\right)^{\left(1-\frac{1}{\beta}\right)} \tag{1}$$

and

$$\frac{\text{MWLT}}{\text{ELT}} = \left(\frac{-(1+f)\ln R_0}{K(\gamma, f)}\right)^{\left(1-\frac{1}{\beta}\right)}. \tag{2}$$

Table 1 displays values of equations (1) and (2) for selected values of β, R_0, f, and $\gamma = .80$. When $\beta < 1$, MWLT as defined by equation (15) is actually the minimum value of WLT and mWLT is the maximum value of WLT. See the Appendix section for details.

TABLE 1

VALUES OF $\left(\frac{\text{MWLT}}{\text{ELT}}, \frac{\text{mWLT}}{\text{ELT}}\right)$ FOR 80% LCL

	$R_0 = .90$			$R_0 = .95$		
β	$f = 1$	$f = 2$	$f = 3$	$f = 1$	$f = 2$	$f = 3$
.8	1.94, 2.31	1.92, 2.52	1.90, 2.69	2.32, 2.76	2.30, 3.02	2.28, 3.22
.9	1.34, 1.45	1.34, 1.51	1.33, 1.55	1.45, 1.57	1.45, 1.63	1.44, 1.68
1.1	0.79, 0.74	0.79, 0.71	0.79, 0.70	0.74, 0.69	0.74, 0.67	0.74, 0.65
1.2	0.64, 0.57	0.65, 0.54	0.65, 0.52	0.57, 0.51	0.57, 0.48	0.58, 0.46
1.25	0.59, 0.51	0.59, 0.48	0.60, 0.43	0.51, 0.44	0.51, 0.41	0.52, 0.39
1.3	0.54, 0.46	0.55, 0.43	0.55, 0.40	0.46, 0.39	0.46, 0.36	0.47, 0.34
1.5	0.41, 0.33	0.42, 0.29	0.42, 0.27	0.32, 0.26	0.33, 0.23	0.33, 0.21
2.0	0.27, 0.19	0.27, 0.16	0.28, 0.14	0.19, 0.13	0.19, 0.11	0.19, 0.10

In Table 1 for $R_0 = .90$, $f = 1$, $\beta = 1.25$ the ratio MWLT/ELT is 0.59. That is if time to failure has a Wei(t; λ, 1.25) distribution and $\beta = 1.25$ is used to compute T_0^{β} in the time truncated test, then the largest linear test time required to demonstrate an 80% LCL of .90 for $R(t_0)$ is roughly 0.59 of the test time required to demonstrate this same LCL specification if we assume the time to failure has an exponential distribution. This is a significant reduction in required linear test time for such a small amount of wearout; i.e., $\beta = 1.25$. Note that this ratio is the same for all values of t_0.

APPENDIX

If f failures occur during the sequential time truncated test plan then

$$\sum_{i=1}^{f+1} T_i^{\beta} = T_0^{\beta}. \tag{3}$$

It is well known, [1], that if T is $\varepsilon(\lambda)$ and time censored testing is performed as described in the Introduction section ($\beta = 1$), then the $100\gamma\%$ LCL, $R(t_0)_L$, for $R(t_0)$ is

$$R(t_0)_L = e^{-\lambda_u t_0} \tag{4}$$

where $\lambda_u = K(\gamma, f)/T_0$. Setting the right number of equation (4) equal to R_0 and solving for T_0 yields

$$T_0 = \frac{t_0\, K(\gamma, f)}{-\ln R_0} \equiv \text{ELT}. \tag{5}$$

If T is Wei(t; λ, β), $\beta \neq 1$, then T^{β} is $\varepsilon(\lambda^{\beta})$. In this case, the $100\gamma\%$ LCL, $R(t_0)_L$, for $R(t_0) = e^{-(\lambda t)^{\beta}}$ is

$$R(t_0)_L = e^{-\left(\lambda^{\beta}\right)_u t_0^{\beta}} \tag{6}$$

when $\left(\lambda^{\beta}\right)_u = K(\gamma, f)\big/T_0^{\beta}$. Setting the right member of equation (6) equal to R_0 and solving for T_0 yields

$$T_0 = t_0\left(\frac{K(\gamma, f)}{-\ln R_0}\right)^{\frac{1}{\beta}}. \tag{7}$$

When $\beta \neq 1$, WLT is random. When f failures occur,

$$\text{WLT} = \sum_{i=1}^{f} T_i + T_{f+1} \tag{8}$$

and from equation (3)

$$T_{f+1} = \left(T_0^{\beta} - \sum_{1}^{f} T_i^{\beta}\right)^{\frac{1}{\beta}}. \tag{9}$$

Consequently WLT is a function of $T_1, T_2, \ldots, T_f$. If $\beta > 1$, the maximum of WLT occurs at values of $T_1, T_2, \ldots, T_f$ that satisfy the set of f equations

$$\frac{\partial}{\partial T_i}\left(\sum_{1}^{K} T_i + \left(T_0 - \sum_{1}^{f} T_1^{\beta}\right)^{\frac{1}{\beta}}\right) = 0 \qquad i = 1,2,\ldots,f. \tag{10}$$

Taking partial derivatives yields the f equations

$$1-T_i^{\beta-1}\left(T_0-\sum_1^f T_1^{\beta}\right)^{\frac{1}{\beta}-1}=0 \qquad i=1,2,\ldots,f. \tag{11}$$

Solving for T_i yields

$$T_i=\left(T_0-\sum T_1^{\beta}\right)^{\frac{1}{\beta}}=T_{f+1} \qquad i=1,2,\ldots,f. \tag{12}$$

That is, the maximum of WLT occurs when all T_i are equal and

$$(f+1)T_i^{\beta}=T_0^{\beta}. \tag{13}$$

Consequently,

$$T_i=T_0\Big/(f+1)^{\frac{1}{\beta}} \qquad i=1,2,\ldots,f+1. \tag{14}$$

Therefore,

$$\text{MWLT}=\sum_{i=1}^{f+1}T_i=T_0(f+1)^{1-\frac{1}{\beta}} \tag{15}$$

where T_0 is given in equation (7). Dividing equation (15) by equation (5) with T_0 replaced by its expression in equation (7), yields

$$\frac{\text{MWLT}}{\text{ELT}}=\left(\frac{-(f+1)\ln R_0}{K(\gamma,f)}\right)^{1-\frac{1}{\beta}}. \tag{16}$$

The smallest value of WLT, mWLT, will be equal to the T_0 of equation (7) which occurs if $T_1 = T_0$. That is, mWLT = T_0. Dividing equation (7) by equation (5) yields

$$\frac{\text{mWLT}}{\text{ELT}}=\left(\frac{-\ln R_0}{K(\gamma,f)}\right)^{1-\frac{1}{\beta}}. \tag{17}$$

If $\beta < 1$, the expression for MWLT in equation (15) yields the minimum value of WLT in equation (8) and T_0 is its maximum value. That is, MWLT is the minimum value of WLT and mWLT is the maximum value of WLT.

REFERENCES

1. Bain, L.J., and M. Engelhardt, Introduction to Probability and Mathematical Statistics, 2nd edition, Duxbury Press, 1992.

A Bayesian Pareto Analysis for System Optimization

James R. Thompson and Roxy D. Walsh, Rice University

1 Introduction

We present a brief outline for a strategy for system optimization for the United States Space Station. Most statistical derivations have been relegated to the Statistical Appendix, which the reader may choose to reference as necessary. A more detailed outline for statistics and probability may be found in the appendix of [8].

There are a number of paradigms for quality improvement. Our orientation will be toward that of the late W. Edwards Deming [1,3]. Nevertheless, since our special interest here concerns techniques for system optimization of the NASA Space Station, we will find it appropriate to make some modifications in order to account for the special challenges which the Space Station presents.

Deming was largely concerned with industrial systems which had produced numerous rather similar goods over long periods of time. The Space Station is unique, and it has yet to be built. When it has been completed, it will be, for some time, a "one of a kind" system, highly complex, new in concept and purpose, without the luxury of direct experiential information.

There is always the temptation when dealing with a new project as revolutionary as the Space Station surely is to "start from zero" to assume that one is forced to deal with the completely new. In our opinion, such a temptation should be resisted. There are, admittedly, less complex systems which can shed light on the task of system optimization of the Space Station. To achieve integration of such information is a formidable task, and in such a short time we can only attempt to formulate a framework by which this task might be achieved.

2 Pareto's Maxim

The philosophy of Deming is based on ancient precedents. First, there are the notions of logical consistency and the reproducibility of experiments. These discoveries of Aristotle, buttressed by St. Paul and St. Thomas Aquinas, clearly most important in the ethos of the West, form the basis for the so-called "scientific method," with which the Deming paradigm is completely consistent. Then, there is a harking back to the harmony of the late Middle Ages, when craft guilds in the cities of Europe formed the early modalities of production, based not so much on laissez-faire competition as on cooperation. Deming had very little patience with decisions based on short range economic gain. Suppliers should not be changed lightly, for a change in input to a system must generally produce changes in the output of the system. In this regard, it should be noted, Deming follows the lead of Henry Ford [2], whose empirical adherence to something very like SPC, put him at variance with classical freetraders. And, like the medieval masters of crafts, Deming holds as sacred the encouragement and skill development of workers. To Deming, throwing away an experienced and well trained employee is not only wicked but stupid, at least as stupid as throwing away

money.

Most important as a precursor of Deming was the Italian sociologist and economist (his doctorate, however, being in civil engineering) Vilfredo Pareto. A major basis of the paradigm of statistical process control is the empirical observation of Vilfredo Pareto that *the failures in a system are usually the consequence of a few assignable causes rather than the consequence of a general malaise across the system.* This insight is generally known as *Pareto's Maxim.* It is, of course, easy to state "laws" and "maxims". There is little reason, *a priori*, to give Pareto's Maxim any more credence than Pyramid Power or Transcendental Meditation. Perhaps the greatest of the considerable accomplishments of W. Edwards Deming is the demonstration of nearly half a century as to the practical implication of the Maxim of Pareto.

To give an example of Pareto's Maxim, let us imagine a room filled with blindfolded people which we would wish to be quiet but is not because of the presence of a number of noise sources. Most of the people in the room are sitting quietly, and contribute only the sounds of their breathing to the noisiness of the room. One individual, however, is firing a machine gun filled with blanks, another is playing a portable radio at full blast, still another is shouting across the room, and, finally, one individual is whispering to the person next to him. Assume that the "director of noise diminution" is blindfolded also. Any attempt to arrange for a quiet room by asking everyone in the room to cut down his noise level 20% would, of course, be ridiculous. The vast majority of the people in the room, who are not engaged in any of the four noise making activities listed, will be annoyed to hear that their breathing noises must be cut 20%. They rightly and intuitively perceive that such a step is unlikely to do any measurable good. Each of the noise sources listed is so much louder than the next down the list that we could not hope to hear, for example, the person shouting until the firing of blanks had stopped and the radio had been turned off.

The prudent noise diminution course is to attack the problems sequentially. We first get the person firing blanks to cease. Then, we will be able to hear the loud radio, which we arrange to have cut off. Next, we can hear the shouter, request that he be quiet. Finally, we can hear the whisperer and request that he also stop making noise.

If we further have some extraordinary demands for silence, we could begin to seek the breather with the most clogged nasal passages, and so on. But generally speaking, we would arrive, sooner or later, at some level of silence which would be acceptable for our purposes. This intuitively obvious analogy is a simple example of the key notion of quality control. By standards of human psychology, the example is also rather bizarre. Of the noise making individuals, at least two would be deemed sociopathic. We are familiar with the fact that in most gatherings, there will be a kind of uniform buzz. If there is a desire of a master of ceremonies to quiet the audience, it is perfectly reasonable for him to ask everyone please to be quiet. The fact is that machines and other systems tend to function like the (by human standards) bizarre example and seldom behave like a crowd of civilized human beings. It is our tendency to anthropomorphize systems that makes the effectiveness of statistical process control appear so magical.

Following the Maxim of Pareto, the basic approach of Deming is to prioritize the investigations of potential causes of system suboptimality so that **we spend our resources on dealing with those where difficulties are most likely to be found**. The general

device used by Deming to achieve this prioritization is the Control Chart largely based on mean behavior. In mature systems operating for a reasonable time, the mean control chart will prove invaluable. However, as we shall see later on, we need to make some modification in order to handle a new, one of a kind system such as the Space Station. The approach we shall develop will be in the spirit of Deming, though our techniques will be somewhat modified to handle the "one of a kind" situation presented by the Space Station.

3 Deming's Basic Approach

In our preliminary material, which is oriented toward industrial production, we rely heavily on material from books by Thompson [4], [6], Thompson and Koronacki [5] and Williams and Thompson [7]. Problems encountered in the optimization, say, of the Space Station are somewhat different from those of industrial production, since they involve a fundamentally new and unduplicated system. The wealth of data assumed by the standard techniques of Statistical Process Control is not to be taken as a given. Hence, we shall develop a procedure which allows much more utilization of historical analogy and expert opinion than is characteristic of the Deming paradigm. Nevertheless, no discussion of system optimization would be complete without attention to the management paradigm of the late W. Edwards Deming. And our treatment will be very much in the spirit of Deming with modifications necessary for the highly complex "one of a kind" system represented by the Space Station.

At the end of the Second World War, Japan was renowned for shoddy goods produced by automatons living in standards of wretchedness and resignation. The formalism of the Zen culture of Japan appeared to be at the opposite pole of Aristotelian realism which characterized the nations of the First World. If there were ever a society apparently unpromising for rapid industrial progress, amongst the civilized nations of the world, postwar Japan would appear to appear to have been amongst the most unpromising.

Deming began preaching his paradigm of Statistical Process Control in Japan in the early 1950s. By the mid 1960s, Japan was a serious player in electronics and automobiles. By the 1980s, Japan had taken a dominant position in consumer electronics and, absent tariffs, automobiles. Even in the most sophisticated areas of production, for example, computing, the Japanese had achieved a leadership role. The current situation of the Japanese workers is among the best in the world. A miracle, to be sure, and one far beyond that of, say, postwar Germany, which was a serious contender in all levels of production before World War II.

It would seem impossible that the Deming paradigm, which involves no new hardware and which, culturally, seems poles apart from Zen formalism and notions of group tranquility, could have made the difference. The fact is that Deming had made several incredibly important observations [7]:

- The key to optimizing the output of a system is the optimization of the system itself.
- Although the problem of modifying the output of a system is frequently one of linear feedback(easy), the problem of optimizing the system itself is one of nonlinear feedback(hard).

- The suboptimalities of a system are frequently caused by a few assignable causes. These manifest themselves by intermittent departures of the output from the overall output averages.
- Hence, it is appropriate to dispense with complex methods of system optimization and replace these by human intervention whenever one of these departures is noted.
- Once an assignable cause of suboptimality has been removed, it seldom recurs.
- Thus, we have the indication of an apparently unsophisticated but, in fact, incredibly effective paradigm of system optimization.

4 Pareto and Ishikawa Diagrams

In free market economies, we are in a different situation than managers in command economies, for there generally is a "bottom line" in terms of profits. The CEO of an automobile company, for example, will need to explain the dividends paid per dollar value of stock. If it turns out that these dividends are not satisfactory, then he can take "dramatic action" such as having his teams of lobbyists demand higher tariffs on foreign automobiles and instructing his advertising department to launch intimidating "buy American" campaigns. Sometimes, he might take even more dramatic action by trying to build better automobiles (but that is unusual). We note that if the decision is made to improve the quality of his product then there is the question of defining what it means for one car to be better than another. It is all very well to say that if profits are good, then we probably are doing OK, but a reasonable manager should look to the reasons why his sales should or should not be expected to rise. Uniformity of product is the measure which we will be using to a very large degree in the development of the statistical process control paradigm. But clearly, this is not the whole story. For example, if a manufacturer was turning out automobiles which had the property that they all ran splendidly for 10,000 miles and then the brake system failed, that really would not be satisfactory as an ultimate end result, even though the uniformity was high. But, as we shall see, such a car design might be very close to good if we were able simply to make appropriate modification of the braking system. A fleet of cars which had an average time to major problems of 10,000 miles but with a wide variety of failure reasons and a large variability of time til failure would usually be more difficult to put right.

The modern automobile is a complex system with tens of thousands of basic parts. As with most real world problems, a good product is distinguished from a bad one according to an implicit criterion function of high dimensionality. A good car has a reasonable price, "looks good," has good fuel efficiency, provides safety for riders in the event of an accident, has comfortable seating in both front and rear seats, has low noise levels, reliably starts without mishap, etc., etc.

Yet, somehow, consumers manage to distill all this information into a decision as to which car to purchase. Certain criteria seem to be more important than others. For example, market analysts for years have noted that Japanese automobiles seem to owe their edge in

large measure to the long periods between major repair. One hears statements such as, "I just changed the oil and filter every five thousand miles, and the thing drove without any problems for 150,000 miles."

Long time intervals between major repairs make up one very important criterion with American car buyers. Fine. So then, an automotive CEO might simply decide that he will increase his market share by making his cars have long times til major repairs. How to accomplish this? First of all, it should be noted that broad spectrum pep talks are of negative utility. Few things are more discouraging to workers than being told that the company has a problem and it is up to them to solve it without any clue as to how this is to be achieved.

A reasonable first step for the CEO would be to examine the relative frequencies of causes of first major repair during a period of, say, three months. The taxonomy of possible causes must first be broken down into the fifty or so groups. We show in Figure 1 only the top five. It is fairly clear that management needs to direct a good deal of its attention to improving transmissions. Clearly, in this case, as is generally true, a few causes of difficulty are dominant. The diagram in Figure 1 is sometimes referred to as a *Pareto diagram*, inasmuch as it is based on *Pareto's Maxim* to the effect that *the failures in a system are usually the consequence of a few assignable causes rather than the consequence of a general malaise across the system.*

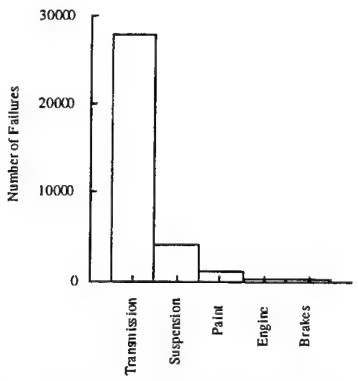


Figure 1. Failure Pareto Diagram.

What is the appropriate action of a manager who has seen Figure 1? At this point, he could call a meeting of the managers in the Transmission Section and tell them to fix the problem. This would not be inappropriate. Certainly, it is much preferable to a general harangue of the entire factory. At least he will not have assigned equal blame to the Engine Section with 203 failures (or the Undercoating Section with no failures) as to the Transmission Section with 27,955 failures. The use of hierarchies is almost inevitable in management. The Pareto diagram tells top management where it is most appropriate to spend resources in finding (and solving) problems. To a large extent, the ball really is in the court of the Transmission Section (though top management would be well advised to pass through the failure information to the Suspension Section and indeed to all the sections).

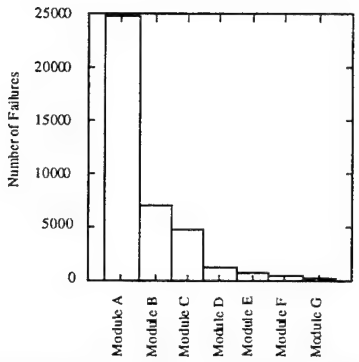


Figure 2. Transmission Failure Pareto Diagram.

What should be the approach of management in the Transmission Section? The obvious answer is a Pareto diagram on the 27,955 faulty transmissions. That may not be realistic. It is easier to know that a transmission has failed than what was the proximate cause of that failure. We might hope that the on site mechanics will have correctly diagnosed the problem. Generally speaking, in order to save time, repair diagnostics will be modularized; i.e., there will be a number of subsections of the transmission which will be tested as to whether they are satisfactory or not. Naturally some of the transmissions will have more than one failed module.

Clearly, Module A is causing a great deal of the trouble. It is possible to carry the hierarchy down still another level to find the main difficulty with that module. The problem may be one of poor design, or poor quality of manufacture. Statistical process control generally addresses itself to the second problem.

The "cause and effect" or "fishbone" diagram of Ishikawa is favored by some as a tool for finding the ultimate cause of a system failure. Let us demonstrate what such a diagram might look like for the present problem.

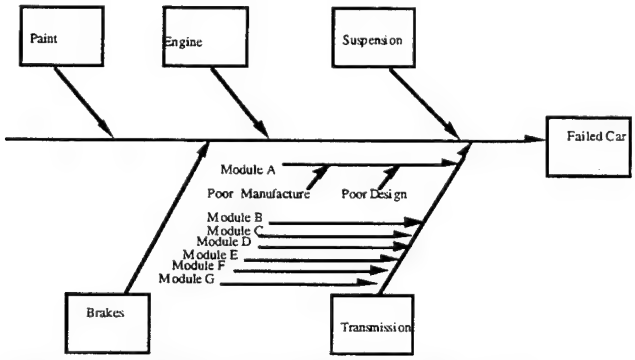


Figure 3. Fishbone (Ishikawa) Diagram.

The fishbone diagram should not be thought of as a precise flowchart of production. The chart as shown might lead one to suppose that the transmission is the last major component

installed in the car. That is not the case. We note that Figure 3 allows for free form expression such as might come about in a discussion where a number of people are making inputs as to an appropriate representation on the blackboard. Each of the paths starting from a box is really a stand-alone entity. We have here developed only one of the paths in detail. We note that in the case of Transmissions, we go down the next level of hierarchy to the modules and then still one more level to the design and quality of manufacturing. In practice, the fishbone diagram will have a number of such paths developed to a high level of hierarchy. Note that each one of the major branches can simply be stuck onto the main stem of the diagram. This enables people in "brainstorming" sessions to submit their candidates for what the problem seems to be by simply sticking a new hierarchy onto the major stem.

5 An Approach for the Space Station

The foregoing industrial examples bear on system optimization for the Space Station. Yet they differ in important aspects. An industrialist might, if he so chooses, simply allocate optimization resources based on customer complaints. We note that we were dealing with nearly 30,000 cases of transmission complaints alone. We have no such leisure when we consider system optimization of the Space Station. We cannot simply wait, calmly, to build up a data base of faulty seals and electrical failures. We must "start running" immediately. Thus, we will require an alternative to a hierarchy of histograms. Yet there are lessons to be learned from the industrial situation.

5.1 Hierarchical Structure

First of all, in the case of building a car, we recall that we had a hierarchy of parts of the system to be optimized. We did not simply string out a list of every part in a car. We formed a hierarchy, in the case of a car, we had three levels. Possibly, in the complexity of the Space Station, we will need to extend the hierarchy to a higher number than three, possibly as high as six or seven levels.

A top level might consist, say, of structure, fluid transmission, life support, electromechanical function, kinetic considerations and data collection. Again, we note that modern quality control seldom replaces a bolt or a washer. The irreducible level is generally a "module." We would expect such a practice to be utilized with the Space Station also. If we assume that we have a hierarchy of six levels and that there are roughly seven sublevels for each, then we will be dealing with approximately $7^6 = 117,649$ basic module types for consideration.

In Figure 4 below, we demonstrate the sort of hierarchical structure we advocate through three levels. Even at three levels, using seven categories at each stage, we would be talking about $7^3 = 343$ end stages.

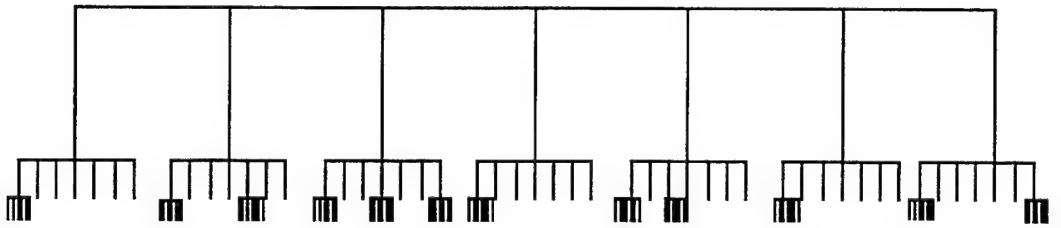


Figure 4.Three Levels of Hierarchy.

5.2 Pareto's Maxim Still Applies

Again, in the case of the Space Station, it would be folly to assume that at each level of the hierarchy, the probability of less than satisfactory performance in each category is equally likely. We do not have experiential histograms to fall back on. Classical flow charting will not be totally satisfactory, at least in the early days of operation. We need an alternative to the (say) six levels of histograms.

5.3 A Bayesian Pareto Model

Let us suppose that at a given level of hierarchy, the failures (by this we mean any departures from specified performance) due to the k components are distributed independently according to a homogeneous Poisson process. So, if t is the time interval under consideration, and the rate of failure of the ith component is θ_i, then the number y_i of failures in category i is given (see C.2.1) by

$$f(y_i|\theta_i) = \exp(-\theta_i t)\frac{(\theta_i t)^{y_i}}{y_i!} \ . \tag{1}$$

The expected number of failures in category i during an epoch of time length t is given by

$$E(y_i|\theta_i) = \sum_{y_i=0}^{\infty} y_i e^{-\theta_i t}\frac{(\theta_i t)^{y_i}}{y_i!} = \theta_i t$$

Similarly, it is an easy matter to show that the variance of the number of failures in category i during an epoch of time length t is also given, in the case of the Poisson process by $\theta_i t$. Prior to the collection of failure data, the distribution of the ith failure rate is given by the *prior density*:

$$p(\theta_i) = \frac{\theta_i^{\alpha_i - 1}\exp(-\frac{\theta_i}{\beta_i})}{\Gamma(\alpha_i)\beta_i^{\alpha_i}} \tag{2}$$

Then, the *joint density* of y_i and θ_i is given by taking the product of $f(y_i|\theta_i)$ and $p(\theta)$:

$$f(y_i, \theta_i) = \exp(-\theta_i t)\frac{(\theta_i t)^{y_i}}{y_i!}\frac{\theta_i^{\alpha_i - 1}\exp(-\frac{\theta_i}{\beta_i})}{\Gamma(\alpha_i){\beta_i}^{\alpha_i}} \tag{3}$$

Then, the *marginal distribution* of y_i is given by

$$\begin{aligned} f(y_i) &= \frac{t^{y_i}}{y_i!\Gamma(\alpha_i)\beta_i^{\alpha_i}} \int_0^\infty \exp(-\theta_i(t+\frac{1}{\beta_i}))\theta_i^{y_i+\alpha_i-1} d\theta_i \\ &= \frac{t^{y_i}}{y_i!\Gamma(\alpha_i)\beta_i^{\alpha_i}(t+1/\beta_i)^{y_i+\alpha_i}}\Gamma(y_i+\alpha_i) \end{aligned} \tag{4}$$

Then the *posterior density* of θ_i given y_i is given by the quotient of $f(y_i, \theta_i)$ divided by $f(y_i)$:

$$g(\theta_i|y_i) = \exp[-\theta_i(t+1/\beta_i)]\theta_i{}^{y_i+\alpha_i-1}(t+1/\beta_i)^{y_i+\alpha_i}/\Gamma(y_i+\alpha_i) \tag{5}$$

Then, looking at all k categories in the level of the hierarchy with which we are currently working, we have for the prior density on the parameters $\theta_1, \theta_2, \ldots, \theta_k$,

$$p(\theta_1, \theta_2, \ldots, \theta_k) = \Pi_{i=1}^k \frac{\theta_i^{\alpha_i-1}\exp(-\frac{\theta_i}{\beta_i})}{\Gamma(\alpha_i)\beta_i{}^{\alpha_i}} \tag{6}$$

Similarly, after we have recorded over the time interval $[0, t]$, $y_1, y_2, \ldots, y_k$ failures in each of the modules at the particular level of hierarchy, we will have the posterior distribution of the θ_i given the y_i,

$$g(\theta_1, \theta_2, \ldots, \theta_k|y_1, y_2, \ldots, y_k) = \Pi_{i=1}^k \exp[-\theta_i(t+1/\beta_i)]\theta_i{}^{y_i+\alpha_i-1}(t+1/\beta_i)^{y_i+\alpha_i}/\Gamma(y_i+\alpha_i) \tag{7}$$

It should be observed in (7) that our prior assumptions concerning α had roughly the same effect as adding α_i failures at the beginning of the observation period.

We note that

$$E[\theta_i] = (y_i+\alpha_i)\frac{t}{t+1/\beta_i} \tag{8}$$

Furthermore,

$$Var[\theta_i] = (\frac{t}{t+1/\beta_i})^2(y_i+\alpha_i) \tag{9}$$

We note that if we rank the expectations from largest to smallest, we may track for each time period, we may plot $E[t\theta_i]$ values to obtain a Bayesian Pareto plot very similar to the Pareto plot in Figure 1.

How shall one utilize expert opinion to obtain reasonable values of the α_i and β_i? First of all, we note that equations (8) and (9) have two unknowns. We are very likely to be able to ask an expert the question, "how many failures do you expect in a time interval of length t?" This will give us the left hand side of equation (8). An expression for the variance is generally less clearly dealt with by experts, but there are various ways to obtain nearly equivalent "spread" information. For example, we might ask the expert to give us the number of failures which would be exceeded in a time interval of length t only one time in ten.

5.4 An Example

Let us suppose that at the top level of hierarchy, we have seven subcategories. At the beginning of the study, expert opinion, leads us to believe that for each of the subcategories, the expected "failure rate" per unit time is 2, and the variance is also 2. This gives us, before any data is collected, $\alpha_i = 2$ and $\beta_i = 1$. So, for each of the prior densities on θ_i we have the gamma density shown in Figure 5.

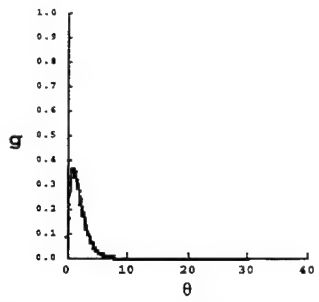


Figure 5. Priors without Data.

However, after 5 time units have passed, we discover that there have been 100 "failures" in the first module, and 5 in each of the other modules. This gives us the posterior distributions shown in Figure 6. Clearly, we now have a clear indication that the posterior on the right (that of the first module) strongly indicates that the major cause of "failures" is in that first module, and that is where resources should be allocated until examination of the evolutionary path of the posteriors in lower levels of the hierarchy give us the clue to the cause of the problem(s) in module seven, which we then can solve.

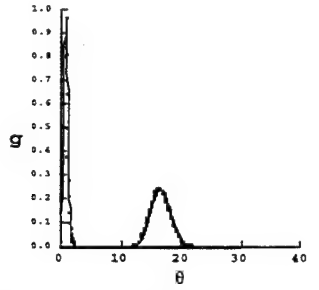


Figure 6. Evolving Posterior Distributions.

Perhaps of more practical use to most users would be a *Bayesian Pareto Chart*, which is simply the expected number of failures in a time epoch of length seven. From (7) we note that

$$E[t\theta_i] = (y_i + \alpha_i)\frac{t}{t + 1/\beta_i}$$

We show such a chart in Figure 7.

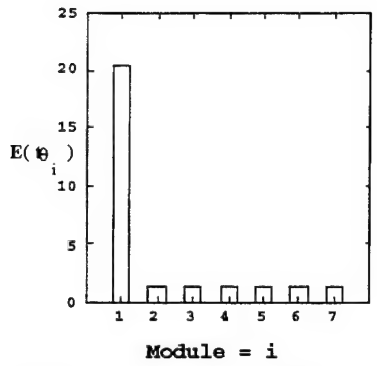


Figure 7. Bayesian Pareto Chart.

One very valid criticism might have to do with the inappropriateness of the assumption that the rates of failure in each category at a given level of hierarchy are independent. The introduction of dependency in the prior density will not be addressed here, since the study of the independent case allows us conveniently to address the evolution of posterior densities without unnecessarily venturing into a realm of algebraic complexity.

5.5 Allowing for the Effect of Elimination of a Problem

It should be noted that when we solve a problem, it is probably unwise to include all the past observations which include data before the problem was rectified. For example, if we fix the first module in Figure 6, then we should discount, in a convenient way, observations which existed prior to the "fix." On the other hand, we need to recognize the possibility that we have not actually repaired the first module. It might be unwise immediately completely to discount those 100 failures in the 5 time units until we are really sure that the problem has been rectified. Even if we did not discount the failures from the time period before the problem has been rectified, eventually the posterior distribution would reflect the fact that less attention needs be given to repairs in the seventh module. But "eventually" might be a long time.

One way to discount records from the remote past is to use an *exponential smoother* such as

$$\hat{z}_i = (1 - r)\hat{z}_{i-1} + rz_i$$

where a typical value for r is 0.25. Let us consider the data in Table 1. Here, a malfunction in the first module was discovered and repaired at the end of the fifth time period. z_{ij} represents the number of failures of the ith module in the tth time period. $z_{i0} = \alpha_i$.

Table 1											
Module	z_{i0}	z_{i1}	z_{i2}	z_{i3}	z_{i4}	z_{i5}	z_{i6}	z_{i7}	z_{i8}	z_{i9}	z_{i10}
1	2	20	18	23	24	15	2	2	0	2	1
2	2	1	1	2	0	1	2	0	1	1	2
3	2	1	2	0	1	1	1	2	1	0	0
4	2	0	2	0	2	1	1	1	1	2	0
5	2	2	1	0	0	2	0	0	2	2	1
6	2	0	2	1	1	1	1	1	1	2	1
7	2	1	1	0	2	1	1	1	1	0	2

Application of the exponential smoother with r=.25 gives the values in Table 2.

Table 2											
Module	$\hat{z}_{i0}$	$\hat{z}_{i1}$	$\hat{z}_{i2}$	$\hat{z}_{i3}$	$\hat{z}_{i4}$	$\hat{z}_{i5}$	$\hat{z}_{i6}$	$\hat{z}_{i7}$	$\hat{z}_{i8}$	$\hat{z}_{i9}$	$\hat{z}_{i10}$
1	2	6.5	9.38	12.78	15.59	15.44	12.08	9.56	7.17	3.29	1.57
2	2	1.75	1.56	1.67	1.25	1.19	1.39	1.04	1.03	1.01	1.75
3	2	1.75	1.81	1.36	1.27	1.20	1.15	1.36	1.27	0.32	0.08
4	2	1.5	1.62	1.22	1.41	1.31	1.23	1.17	1.13	1.78	0.45
5	2	2.00	1.75	1.31	0.98	1.24	0.93	0.70	1.02	1.76	1.19
6	2	1.5	1.62	1.47	1.35	1.26	1.20	1.15	1.11	1.78	1.19
7	2	1.75	1.56	1.17	1.38	1.28	1.21	1.16	1.12	0.28	1.57

In Figure 8, we show the exponentially weighted Pareto chart at the end of time interval 5 and at the exponentially weighted Pareto chart at the end of time interval 10.

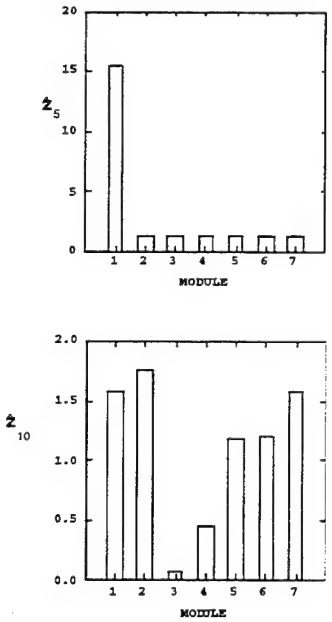


Figure 8. Exponentially Weighted Pareto Charts.

In Figure 9, we show time lapsed exponetially weighted charts for all ten time intervals. It is clear that by the end of the ninth time interval, we should consider relegating module one to a lower level of risk of failures and reallocating inspection resources accordingly.

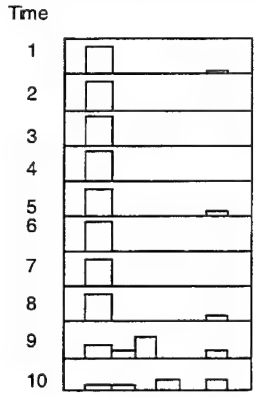


Figure 9. Time Lapsed Exponentially Weighted Pareto Charts.

Acknowledgements

This research was supported in part by the National Aeronautics and Space Administration (Houston) under (NAG 9-771) and the Army Reasearch Office (Durham) under (ARO-DAAH04-95-1-0665).

References

[1] Deming, W. E. (1982). *Quality, Productivity and Competitive Position.* Center for Advanced Engineering Studies, pp. 16-17.

[2] Ford, H. (1926). *My Life and Work.* Sydney: Cornstalk Press, p. 273.

[3] Mann, N. R. (1985). *The Keys to Excellence: the Story of the Deming Philosophy,* Los Angeles: Prestwick Books.

[4] Thompson, J. R. (1985). "American quality control: what went wrong? What can we do to fix it?" *Proceedings of the 1985 Conference on Applied Analysis in Aerospace, Industry and Medical Sciences,* Chhikara, Raj, ed., Houston: University of Houston, pp. 247-255.

[5] Thompson, J.R. and Koronacki, J. (1993). *Statistical Process Control for Quality Improvement,* New York: Chapman & Hall.

[6] Thompson, J.R. (1989). *Empirical Model Building,* New York: John Wiley & Sons.

[7] Williams, E.E. and Thompson, J.R. (1995) *The Economics of Production and Productivity: A Modeling Approach.* Austin: Capital.

INTENTIONALLY LEFT BLANK.

DISCERNING A FAIR FIGHT ON DISSIMILAR SIMULATORS

Lorrie L. Hoffman
Department of Statistics, University of Central Florida
Orlando, Florida 32816-2370

ABSTRACT

We have developed a test statistic via the likelihood ratio approach to test H_0: $p_0 = p_0'$ where p_0 is a parameter in the product of a trinomial with parameters n, p_x, p_y and a binomial with parameters N, $p^* = p_0 p_x + p_y$. We have done so only in the special case where the observed cell frequencies x and y in the trinomial are such that $x + y = n$. This technique is applied to an example comparing warriors on simulators. The requirement in the war games example is to establish the superiority of one player over the other despite the fact that they are engaging in combat environments with unintentional handicaps.

INTRODUCTION

The application presents itself as follows: Two warriors engage in a series of contacts via two simulators. These simulators, for various reasons (software or resolution differences) may exhibit unalike environment portrayals. The goal is to determine whether these warriors are equally skilled. There is a true underlying probability (denoted by p_0) which on a fight contact is the chance that warrior 2 will kill warrior 1. This kill probability is for a realistically (matching) display of the environment.

An initial sampling of *n* gridded sites of the paired simulators' environments is conducted. This is done to classify the *n* sites. There are a number of sites where warrior 1 and warrior 2 see a realistic portrayal of the environment characteristic (a bush, a valley, etc.) , a few sites where warrior 2 does not, some where 2 has it but 1 does not, and sites where the item is missing from both screens. The first and last set of points are classified as matching sites and are denoted by x. When warrior 1 sees the environment characteristic and the opponent does not, this is referred to as an advantage warrior 2 point. It is advantageous to warrior 2 due to the manner of operation of the software in that warrior 1 would take refuge behind the bush but warrior 2 would note a clear shot to be taken. It is assumed that warrior 2 takes the utmost opportunity of this situation and kills warrior 1. The frequency of these sites is denoted by y. Advantage warrior 1 points then number n - x - y.

Now, engagement occurs. The type (match, advantage 2, or advantage 1) of site is not recorded during the sample of N contacts. We assume independence contact to contact. Let t be the number of wins out of the N tries. Associate this value with warrior 2. In the simulated environment warrior 2 does not have probability p_0 of securing a kill at each opportunity (contact). In fact, let us assume that if the site is of the advantage warrior 2 type then the probability of a kill is 1 for warrior 2 and if it is an advantage warrior 1 type point then the probability is 0. Only on matching points is the probability of a kill equal to p_0, the unknown value which we need to test.

More elaborate descriptions of war game simulators can be found in publicly released documents such as the Army RD & A Bulletin.[1,2]

The specific formulation of the problem presents itself as a product of a trinomial and a binomial where the parameter of the binomial is a function of those of the trinomial. Let x, y, n - x - y denote observed cell frequencies in a sample of n observations from a trinomial distribution with probabilities p_x, p_y, $1 - p_x - p_y$. Let t and N - t denote observed cell frequencies in a sample of N observations from a binomial with probabilities p^* and $1 - p^*$ where $p^* = p_0 p_x + p_y$. So we let T = 1 if a site is of the matching type and warrior 2 wins or if the site is of the advantage warrior 2 type. Then Pr (T = 1) = Pr (match and warrior 2 wins) + Pr (advantage warrior 2) = Pr (match) * Pr (warrior 2 wins) + Pr (advantage warrior 2) = $p_0 p_x + p_y$ assuming independence of the environment portrayal and of winning. Thus the frequency t mentioned above is $\sum_{i=1}^{n} T_i$ where T_i is Bernouilli with parameter $p_0 p_x + p_y$.

Approved for public release, distribution is unlimited.

So the probability function is:

$$f(n,x,y,N,t) = \frac{n!N!}{x!y!(n-x-y)!t!(N-t)!} p_x^x p_y^y (1-p_x-p_y)^{n-x-y} (p_0 p_x + p_y)^t (1-p_0 p_x - p_y)^{N-t} \qquad (1)$$

where $0 \le x \le n$, $0 \le y \le n$, $0 \le t \le N$, $0 \le p_x + p_y \le 1$, $0 \le p_0 \le 1$.
The problem is to determine a test for H_0: $p_0 = p_0'$ versus the alternative H_1: $p_0 \neq p_0'$, where p_0 is a specified proportion.

This formulation is more complex than others treated elsewhere in the literature. This is even true when we made the additional simplifying assumption that $n - x - y = 0$. In terms of the simulator application this amounts to stating that the data shows that warrior 2 has all the advantage points. This is not an unrealistic assumption. In a simulation environment it is frequently the case than one piece of hardware has a higher resolution than the other and thus would paint all the bushes, valleys etc. and its paired equipment will portray virtually none.

Quesenberry and Hurst [3] developed methods to deal with simultaneous estimation for multinomial proportions but from a single distribution, not for a product as we have here. Bailey [4] and Fitzpatrick and Scott [5] offer improvements and updates. Goodman [6, 7] does view several multinomial populations but his estimation procedures require the use of contrasts with known coefficients. Our $p_0 p_x + p_y$ is not in this category. Madansky [8] is motivated by an application in the field of reliability and does deal with the product of distributions (but only binomials) and none of his separate binomial parameters are expressed as functions of the other binomial parameters. Koyak[9] is inspired by a common marketing index to view estimation of a function which is the sum of the squared parameter values of a single multinomial distribution. This function is more complicated than a contrast but yet only deals in the single multinomial case. Madansky's use of the likelihood ratio test gave impetus to apply that technique to our problem. This approach is employed in the following sections in order to develop a test statistic. Recently, there has been advocacy for using likelihood ratio tests (see Meeker and Escobar[10]).

Elsewhere we have proposed an alternative conceptualization of this problem[11]. That method uses conservative confidence intervals in a two-stage approach. The first stage focuses on defining the range of possibilities concerning the differences in the "look" of the environments. The second stage addresses the quantification of the battle portion. We are able to attach a 90% confidence level to a range of proportions indicating kill ratio potential on a completely matched (fair) environment. If that range includes $p = .5$ then either warrior could be triumphant in a non-handicapped situation. The advantage of the confidence interval method is that t can be 0 or N and x can be 0 or n, plus there is no requirement for $n - x - y = 0$ which are assumptions used in this paper. The disadvantage is the lack of precision in the intervals.

METHODOLOGY

Presently, we have restricted attention to the case where $n - x - y = 0$. With this consideration, view the probability function defined in (1) as a likelihood function where p_x, p_y and p_0 are now variables. Under the null hypothesis set p_0 to a value and maximize (ignoring constants)

$$L(p_x, p_y) = (p_0 p_x + p_y)^t (1 - p_0 p_x - p_y)^{N-t} p_x^x p_y^{n-x} \qquad (2)$$

where $0 \le p_x \le 1$, $0 \le p_y \le 1$ and $0 \le p_x + p_y \le 1$.

The assumption made here is that sampling has occurred from both the trinomial and binomial distributions (that is n, $N \neq 0$). In the case where $1 \le x \le n - 1$ and $1 \le t \le N - 1$ this surface defined in (2) is readily investigated. Only the resulting theorems are presented here.

Theorem 1. With $n, N \neq 0$; $1 \le x \le n - 1$ and $1 \le t \le N - 1$ and $p_0 > [((n(N - t)) / (x(n + t))) + 1]^{-1}$ (3)
then $L(p_x, p_y)$ has exactly 1 local maximum interior to the (0,0) - (0,1) - (1,1) triangle at
$p_x = (x/p_0 n)((n + t) / (N + n))$, $p_y = ((n - x) / n)((n + t) / (N + n))$. (4)
If p_0 is smaller than or equal to expression (3) then there is no local maximum interior to the triangle.

Theorem 2. With $n, N \neq 0$; $1 \le x \le n - 1$ and $1 \le t \le N - 1$ then $L(p_x, p_y)$ has 1 and only 1 local maximum along the wall $p_x + p_y = 1$. The maximum occurs at $p_x = 1 + \Gamma - (\Gamma^2 + ((p_0 (n - x)) / (1 - p_0)(N + n)))^{1/2}$
where $\Gamma = \{p_0 N + n(2p_0 - 1) + (1 - p_0)x - t\} / \{2(1 - p_0)(N + n)\}$ (5)
and thus $p_y = 1 - p_x$. If $p_0 = 1$ there is no local maximum along the wall $p_x + p_y = 1$.

Theorem 3. With n, N≠ 0; $1 \le x \le n-1$ and $1 \le t \le N-1$ and $x/n \ge 1 - (t/N)$ then $L(p_x, p_y, p_0)$ is maximized at
$p_x = x/n$, $p_y = (n-x)/n$, $p_0 = (tn/Nx) - ((n-x)/x))$ (6)
having functional value $((t/N)^t (1 - (t/N))^{N-t} (x/n)^x ((n-x)/n))$.

Theorem 4. With n, $N \neq 0$; $1 \le x \le n-1$ and $1 \le t \le N-1$ and $x/n < 1 - (t/N)$ then the two candidates for maximizing $L(p_x, p_y, p)$ are the points $p_0 = [((n(N-t))/(x(n+t)))+1]^{-1}$,
$p_y = ((n-x)/n)((n+t)/(N+n)))$, $p_x = 1 - p_y$ (7) and
$p_0 = 0$, $p_x = 1 + [(x - (n+t))/(N+n)]$, $p_y = 1 - p_x$ (8)

These theorems were all proved in the standard manner by viewing the partial derivatives with respect to p_x, p_y and p_0 of the likelihood function expressed in (2). For example, in proving Theorem 1 we noted that $\partial L(.)/\partial p_x$ has $N + x + 1$ roots. There are $x - 1$ roots at $p_x = 0$ which give L(.) a value of zero and are of no interest. Also, the $t - 1$ roots at $p_x = -p_y/p_0$ and the $N - t - 1$ at $p_x = (1 - p_y)/p_0$ are of no interest since they lie outside our constraint triangle. The important factor in $\partial L(.)/\partial p_x$ is $-xp_y^2 - (N+x)p_0^2 p_x^2 - (N+2x)p_0 p_x p_y + (x+t)p_0 p_x + xp_y$. Viewed from the perspective that p_y is constant, this is a parabola. We generated those two roots. Details can be found in Hoffman.[12]

Theorem 5. Given a probability function of the form in (1) where the data collected are such that n, $N \neq 0$; $1 \le x \le n-1$ and $1 \le t \le n-1$ then in order to test H_o: $p_0 = p_0'$ versus H_1: $p_0 \neq p_0'$ at $\alpha = \alpha_0$ calculate $c = -2 \ln (L(\theta)/L(\Omega))$ where $L(\theta)$ = maximum of the likelihood function in (2) evaluated at points identified in (4) and (5); $L(\Omega)$ = maximum of the likelihood function in (2) evaluated at points identified in (6) or in (7) and (8). When $c > \chi^2 (df = 1, 1 - \alpha_0)$ we reject H_0: $p_0 = p_0'$.

Theorem 5 is a direct application of Wilkes work.[13]

The preference is to not rely upon asymptotics. But, if we must it is imperative that we understand how good the fit is, particularly in the small sample size cases. We therefore conduct a simulation (discussed below) to ascertain the appropriateness of using the chi-square distribution.

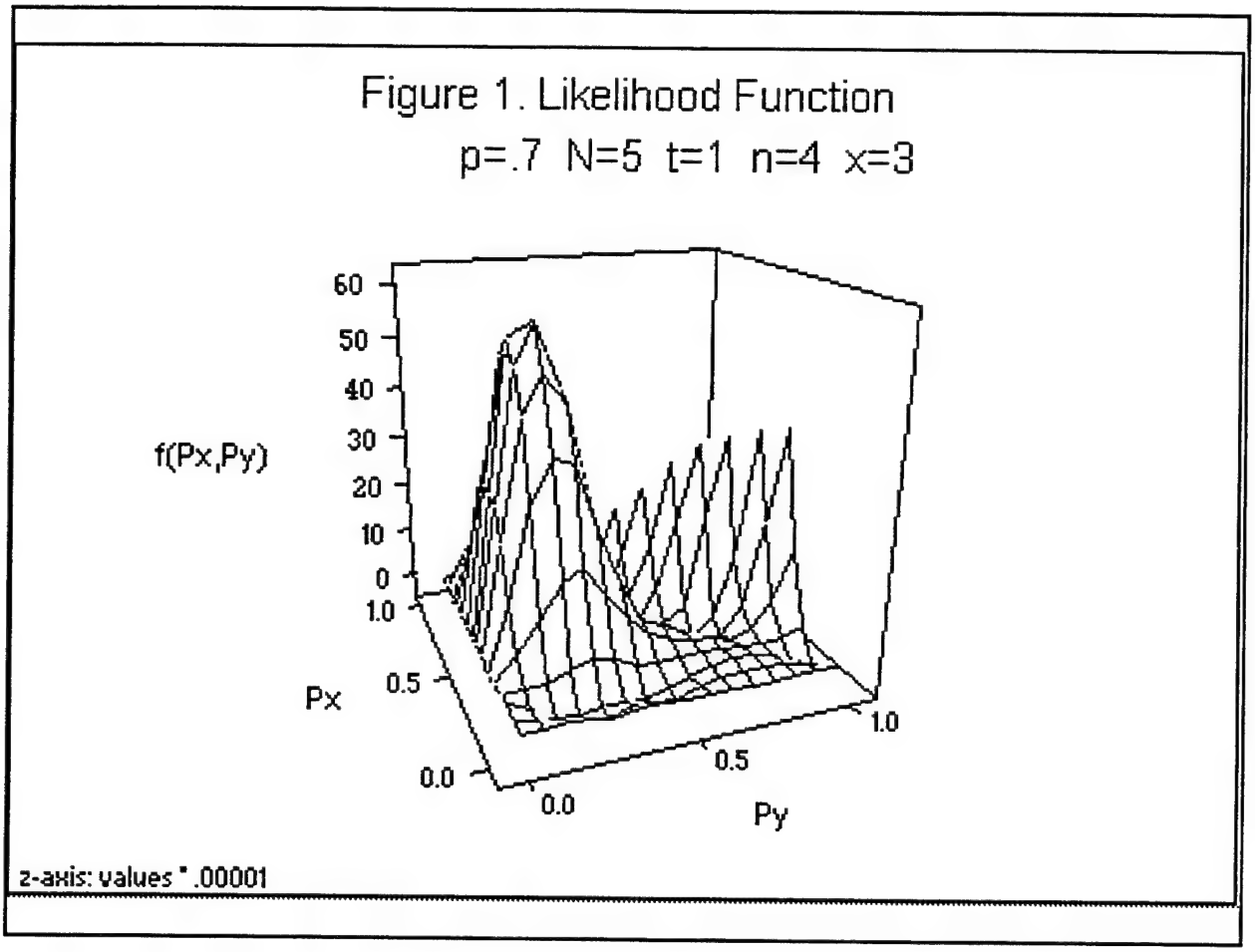
EXAMPLE

With the simulators online we randomly select 4 common points on each of the two screens and note that 3 of them portray the same physical attributes and that 1 point has a bush painted on one screen but not on the other. So we set n=4 and x=3. Next we allow the two warriors to engage in 5 battles. The warrior who has the "inferior" screen (i.e. no bush painted) actually has the advantage (according to the data) and we see that he manages 1 kill in the 5 opportunities. Thus, N=5 and t=1. Let us test H_0:p_0=.7 . That is $p_0 = .7$ means that we believe that our "advantaged" warrior is truly 2.33=.7/.3 times better than his opponent. A quick assessment of this conjecture can be accomplished by considering the following. Seventy-five percent of the points are fair, so .75*5 points where battle occurs, i.e. 3.75 points. If the probability of a kill is .7 then our warrior should win .7*3.75 = 2.625 of those. Additionally, 25% are automatically won by this warrior so credit .25*5 = 1.25 more kills for a total of 2.625+1.25 = 3.875. With the data showing only 1 of 5 we should reject H_0:$p=p_0$=.7.

Here are the formal calculations: with $[((n(N-t))/(x(n+t))+1]^{-1} = .484$ being smaller than p_0=.7 we must evaluate expression (4). We find p_x=.595, p_y=.1338 so $p_0p_x + p_y = .555$ and L(.) = .00063. Expression (5) must always be evaluated. We calculate Γ=.926, $p_x = .869$ with $p_y = .131$ so $p_0p_x + p_y = .739$ and L(.) = .000292. So our $L(\Omega) = .00063$. Figure 1 shows this constrained likelihood surface.

Now we turn to the unconstrained maximum calculations. Since $x/n = 3/4 < 1 - (t/N) = 4/5$ we must compute both expressions (7) and (8). In expression (7) $p_0 = .484$ and $p_y = .1388$ and $p_x = .8611$ so $p_0p_x + p_y = .555$ and L(.) = .00194. For expression (8) we get $p_x = .777$ and $p_y = .223$ and thus L(.)=.0085 . So our $L(\Omega) = .0085$. Using Theorem 5 we calculate c= $-2\ln (L(\theta)/L(\Omega)) = 5.18$. Since $\chi^2 (df=1, 1 - \alpha = .95) = 3.84$ we reject H_0:p_0=.7 .

In this example we were required to calculate both and interior local maximum and a wall local maximum under H_0:p=.7. Also for the unconstrained case since the usual estimator did not apply we were forced to view the two possibilities for maximums given in (7) and (8). Note that the small sample sizes were selected to ease computational burden and that the conclusion may be inaccurate since the statistic c is only asymptotically chi-square. We discuss the asymptotic nature below.



DISCUSSION

We want to understand the risk involved with small samples in using the chi-square as our sampling distribution for our test statistic. We decided to gain insight by viewing some simulation results.

Since we are concerned with the situations where p_z is small (otherwise n - x - y = z can not feasibly be zero), we explored only those cases. We selected the cases where $p_x = .8$ with $p_y = .20, .18, .16$ and $p_x = .85$ with $p_y = .15, .13, .11$ and $p_x = .9$ with $p_y = .10, .08, .06$ and $p_x = .95$ with $p_y = .05, .03, .01$. We considered $p_0 = .1, .3, .5, .7, .9$ for each of the cases along with N and n equal to 5, 12, and 22. This resulted in viewing 180 separate simulations. Since the chi-square with one degree of freedom is proper only where H_0 is true we needed to generate the t and x counts under this assumption. So a "success" occurred (increasing x) when a generated random unit u was less than p_x and likewise t was incremented by 1 when $u < p_0 p_x + p_y$. For each case of the 180 cases we generated 250 pairs of (t,x). We then derived our $-2\ln (L(\theta)/L(\Omega))$ via a grid search for maxima.

We did indeed study all 180 cases, but we only need to present a few to discuss the general conclusions which seem to present themselves. We illustrate these summarizations in Tables 1 , 2 and 3. Those tables exhibited are for certain values of p_0, p_x and p_y and N=n=5, 12, and 22 and contain the observed significance level (p-value) of the simulated distribution corresponding to χ^2 with one degree of freedom critical values of 2.70554 at $\alpha = .10$, 3.84146 at $\alpha = .05$, and 6.6349 at $\alpha = .01$. Additionally, the mean and variance of these simulated distributions are recorded and compared to the mean of 1.0 and variance of 2.0 for the chi-square.

The cases presented in Tables 1 and 2 were the overall worst matches (at all sample sizes) to the expected α-levels. Inherently, this is a difficult testing situation since the simulated binomials with extreme p_x and $p_0\, p_x + p_y$ (since p_0 is .1 and .9) values are highly skewed.

Table 1. Characterization of the distribution of the test statistic when p_0=.10, p_x=.90 and p_y=.16

p-value(.10)	.000	.004	.080
p-value(.05)	.000	.000	.036
p -value(.01)	.000	.000	.000
mean	.720	.810	1.050
variance	.330	.760	1.440
	N=n=5	N=n=12	N=n=22

Table 2. Characterization of the distribution of the test statistic when p_0=.90, p_x=.80 and p_y=.20

p-value(.10)	.000	.008	.164
p-value(.05)	.000	.000	.020
p -value(.01)	.000	.000	.000
mean	.760	.910	1.030
variance	.180	.670	1.640
	N=n=5	N=n=12	N=n=22

In virtually all cases, as the sample size increased the p-values better approximated the chi-square α-levels. It appeared occasionally that the nature of the simulations, ie. random fluctuation, interfered with a statement of this generality. Variability in the α-levels would occur even if the actual distribution were χ^2 with one degree of freedom when conducting n=250 random selections indicating a quantity either above or below the critical value. Here, variations of .02 from the true α-level are usual since $2^* [\alpha *(1 - \alpha) / 250]^{½}$ is approximately .02. So a strict convergence as the sample size increases might not be observed.

The worst case occurred when N=n=5 with p_0=.30, p_x=.95 and p_y=.01. A p-value of .22 resulted. This is in comparison to an expected α-level of .10. This case along with N=n=12 and N=n=22 appears in Table 3. Notice that the asymptotics prevail as the sample size increases.

To summarize, the critical values from a chi-square distribution with one degree of freedom appear (at least from this limited simulation study) to suffice for even relatively small samples sizes of 12. This would mean that for the war games example, we could select 12 points on which to compare the simulator equipment and then later allow the warriors to engage in as few as 12 combat scenarios.

SUMMARY

This is a likelihood ratio test to check on the probabilities of success on a trial conducted on an item drawn from one of the classes of a trinomial population. This is a special case where the data shows no activity from one of the three classes in the trinomial. The construction of the statistic is a straightforward application of the likelihood ratio principle and relies on large sample approximation to compare it to a χ^2 with one degree of freedom which is satisfactory for moderately large (N=n=12) number of trials on the binomial and trinomial. The difficulty in the problem lies in the confounding of the success percentage with the parameters of the trinomial.

Table 3. Characterization of the distribution of the test statistic when p_0=.30, p_x=.95 and p_y=.01

p-value(.10)	.220	.120	.124
p-value(.05)	.024	.052	.062
p -value(.01)	.000	.016	.016
mean	1.270	1.180	1.230
variance	2.330	2.660	2.580
	N=n=5	N=n=12	N=n=22

REFERENCES

1. Vaden, D.W., "Vision for the Next Decade," Army Research Development and Acquisition Bulletin , May - June , p. 1 - 3, 1993.

2. Yuhas, J.S. ,"Distributed Interactive Simulation," Army Research Development and Acquisition Bulletin, May - June, p. 4 - 6,1993.

3. Quesenberry, C.P., and Hurst, D.C. , "Large Sample Simultaneous Confidence Intervals for Multinomial Proportions," Technometrics , vol.6, p. 191 - 195, 1964.

4. Bailey, B.J.R. , "Large Sample Simultaneous Confidence Intervals for the Multinomial Probabilities Based on Transformations of the Cell Frequencies," Technometrics , vol. 22, p. 583 - 589, 1980.

5. Fitzpatrick, S. and Scott, A. "Quick Simultaneous Confidence Intervals for Multinomial Proportions," Journal of the American Statistical Association, vol. 399 , p. 875 - 878, 1987.

6. Goodman, L.A., "Simultaneous Confidence Intervals for Contrasts Among Multinomial Populations," Annals of Mathematical Statistics , vol. 35, p. 716 - 725, 1964.

7. Goodman, L.A.,"On Simultaneous Confidence Intervals for Multinomial Proportions," Technometrics , vol. 7, p. 247 - 254, 1965.

8. Madansky, A.,"Approximate Confidence Limits for the Reliability of Series and Parallel Systems,"Technometrics, vol.7, p. 495 - 503, 1965.

9. Koyak, R., "On Constructing Confidence Intervals for Functions of a Multinomial Parameter," Proceedings of the 23rd Symposium on the Interface, p. 305-308, 1991.

10. Meeker, W.Q., and Escobar, L.A., "Teaching about Approximate Confidence Regions Based on Maximum Likelihood Estimations," The American Statistician, vol. 49, p. 48-53, 1995.

11. Hoffman, L.L.,"The Concept of Fair Fight on Mismatched Simulator Environments,"Proceedings of the Southeastern Simulation Conference, p. 234-236, 1995.

12. Hoffman, L.L.,"Hypothesis Testing on a Product of Dependent Multinomials:A Special Case," University of Central Florida Department of Statistics Technical Report #TR95-37, 1995.

13. Wilkes, S.S., "The Large Sample Distribution of the Likelihood Ratio for Testing Composite Hypotheses," Annals of Mathematical Statitistics, vol. 9, p. 60-62, 1938.

INTRODUCTION TO THE SPECIAL SESSION ON ADVANCED WARFIGHTING EXPERIMENTS (AWEs)

Eugene Dutoit
Dismounted Battlespace Battle Lab
Fort Benning, Georgia 31905

ABSTRACT

This session introduction paper will present the general concept of the Advanced Warfighting Experiment (AWE) and how it fits within the concept of the Battle Labs. These AWEs and the topic of Model-Experiment-Model present statistical implications and problems that should be of interest to the attendees of this conference. A few remarks will be made about the AWE regulations and some insights from experimenters. The papers that were presented in this special session will follow this short introduction.

INTRODUCTION

Battle Labs were formed so that the Army could be responsive to the needs of tomorrow. Today's reality shows that we posses a winning Army of world class excellence. We presently enjoy a technology overmatch when compared to other Armies. However, today's reality also indicates that the Army is undergoing a period of downsizing and will have to continue to operate within an austere resource environment. In contrast, tomorrow's army will be smaller than the present force. Although smaller, the Army of tomorrow will continue to exploit new technologies which will enable it to be more lethal. This increased lethality will allow the Army to carry out missions of global reach and force projection. In order to be responsive to the reduced resources and continued demand to maintain a world class force the Battle Labs are to carry out evaluations and investigations in order to significantly reduce the current acquisition milestones of eight to fifteen years to something much shorter. The goal is to field or acquire systems at a faster rate and reduce technical risks at lower costs.

TRADOC REGULATION 11-1 BATTLEFIELD LABORATORIES PROGRAM

The following statements extracted from the cited TRADOC regulation describe the relationship between the Battle Labs and experimentation.

1. Experimentation with real soldiers and real units is the central work of Battle Labs.

2. Experimentation means discovery learning and listening to soldiers and leaders. Experimental work should lead to requirements.

3. (Experiments) should examine the impacts on doctrine, training, leadership, organization, materiel and the soldier (DTLOMS).

4. Experiments which demonstrate significant added value to warfighting capabilities may result in senior Army leadership decisions for rapid acquisition.

5. The Battle Labs should employ three kinds of simulations; live, constructive and virtual. These are briefly described below.
a. Live simulations employ actual soldiers and equipment operating together. This could happen on instrumented ranges. Operational tests and AWEs are examples of live simulations.

b. Constructive simulations rely on algorithmic and mathematical models. JANUS, CASTFOREM and VIC are examples of constructive simulations.
c. Virtual simulations involve manned simulators interacting within a synthetic environment (i.e., other simulators). SIMNET used for training and development is an example of a virtual simulation.

6. The regulation discusses two kinds of experiments.
a. Advanced Warfighting Experiments are center of gravity culminating efforts focused on a major increase to warfighting capability. They cross many or all of the TRADOC domains of DTLOMS (see paragraph 4 above). Moreover, they impact most, if not all, of the battlefield dynamics and battlefield operating systems. AWEs are approved and prioritized by the CG TRADOC. They have extensive involvement by HQ DA, FORSCOM, AMC and OPTEC.
b. Battle Lab warfighting experiments (BLWEs) are single event or progressive iterative simulations with primary relevance to a single battlefield dynamic. The focus of this special is on the AWE but the BLWE can be considered to be a sub-set of the AWE.

INSIGHTS FROM THE POINT OF VIEW OF AN EXPERIMENTER.

1. Only one organization should be in charge and have control over all aspects of the experiment.

2. Establish clear entry requirements for new systems to participate in the experiment. These entry requirements should be enforced. It may be wise to avoid putting the first prototype into the experiment.

3. Some new systems may require "master operators"; not just basic training. Also, some systems will require new tactics, techniques and procedures (TTPs). Leaders should be trained first to obtain their ideas on new/appropriate TTPs for experimental systems.

4. The integration of multiple systems into the AWE calls for innovative experimental designs versus the traditional single system experiments.

PAPERS THAT WERE PRESENTED IN THIS SESSION.

The following papers were presented in this special session and are given in these Proceedings.

1. McCool, B.; Lyman, J.; Ferguson, J. Evolution of the Model-Test-Model Concept For Use In Operational Testing & Advanced Warfighting Experiments

2. Grynovicki, J. ; Leedom, D. ; Golden, M. ; Wojciechowski, J. Performance Based Metrics for the Digitized Battlefield

Evolution of the Model-Test-Model Concept For Use In Operational Testing & Advanced Warfighting Experiments

Bryson McCool, Jerry Lyman, and LTC John Ferguson
TRADOC Analysis Center - White Sands
White Sands Missile Range, NM 88002

ABSTRACT

Models and simulations (M/S) are valuable tools that can complement and augment live field testing and experimentation (T/E). M/S can be used to develop more comprehensive and cost-effective T/E scenarios and provide valid, credible, and timely operational effectiveness and suitability insights that often cannot be derived directly from the T/E results. In turn, T/E offers the means to calibrate and validate the M/S to ensure battlefield reality is being represented.[1] The Model-Test-Model (MTM) Concept, which embraces these reciprocal relationships between M/S and T/E, was recently used to support the M1A2 Initial Operational Test and Evaluation (IOTE) and the Warrior Focus Advanced Warfighting Experiment (AWE). In each effort, the M/S was used extensively to expand the operational assessment of the T/E while the T/E provided a basis to realistically calibrate the M/S to actual tactical operations.

INTRODUCTION

Appropriate and comprehensive scenario design is critical to the success and utility of force-on-force (FOF) field tests and experiments. If the tests and experiments are well conceived, their execution will occur in an efficient and cost-effective manner with all objectives being met. M/S can provide the T/E scenario development team with additional capability to more effectively and responsively design, develop, and refine operational scenarios that are doctrinally and tactically sound, are robust (i.e., sensitive to the performance of the system of interest), represent an appropriate threat, and can adequately address the critical and pertinent operational T/E issues. Figure 1 graphically depicts the pre-T/E phase methodology.

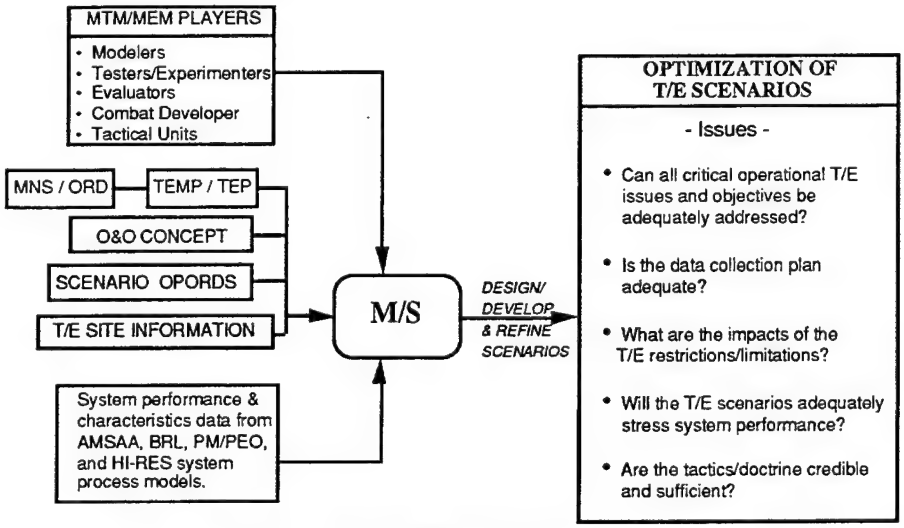


Figure 1. Pre-T/E Modeling and Analysis

The operational orders or OPORDs (which describe the force laydowns, objectives, maneuver unit "avenues of approach," and approximate defensive locations for each of the required scenarios) are normally developed by the appropriate school. The OPORDs and test range terrain data and conditions for each developmental scenario are integrated into the constructive M/S to produce estimates of the force effectiveness measures and battle outcomes that are expected to occur. If the M/S estimates of effectiveness for the 'experimental force' in a particular scenario are not significantly greater than those for the 'baseline' force, an analysis must be performed to determine what changes are required to make the scenario more robust. To reduce T/E costs, constructive M/S can also be used to simulate the baseline case (instead of actually fielding it), the results of which would define force effectiveness if the enhanced or experimental system(s) were not being deployed.

Figure 2 presents an overview of the post-T/E modeling and analysis methodology. After each live FOF event is completed, the player location data for each participating entity, as well as the actual terrain and conditions, are integrated into the constructive simulation. The M/S is then calibrated in an iterative fashion to replicate what occurred in that test or experiment. Calibration is terminated when replication occurs within reasonable tolerances based on statistical testing. The calibrated version of the constructive simulation can be used in several ways to provide valuable insights from the live T/E. T/E issues that could not be adequately addressed during the live experiments (e.g., what would have been the relative contribution of a particular weapon system if obscurants had been used) can be evaluated in a 'what-if' context. In addition, the sensitivity of force effectiveness to specific system performance can be examined to quantify the synergistic impacts of certain critical systems as compared to their collective impact.

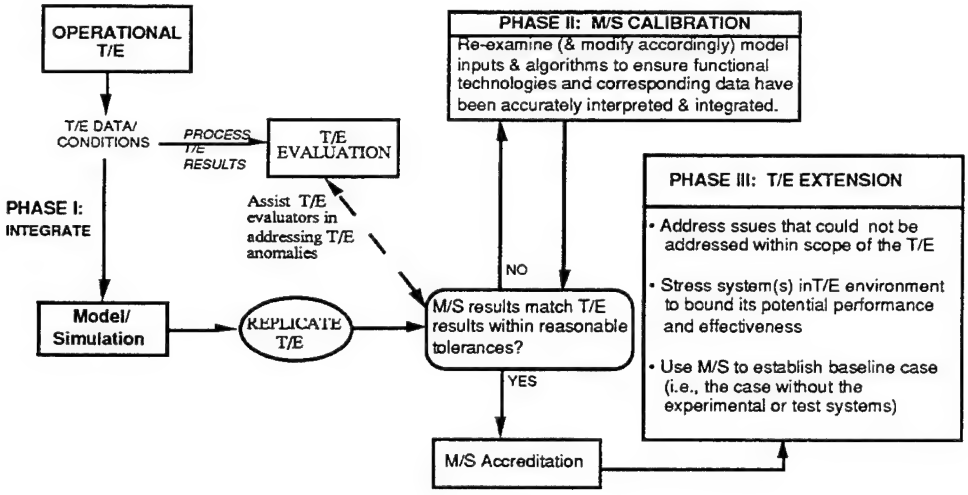


Figure 2. Post-T/E Modeling and Analysis

The TRADOC Analysis Center at White Sands Missile Range (TRAC-WSMR) supported the Operational Test and Evaluation Command (OPTEC) for the M1A2 Initial Operational Test and Evaluation (IOTE). The M1A2 IOTE was performed at Ft. Hood, Texas, by OPTEC from August through December 1993 to assess the operational effectiveness and suitability of the M1A2 Abrams tank under realistic battlefield conditions. The M1A2 Milestone III Cost and Operational Effectiveness Analysis (MS III COEA) was performed by TRAC-WSMR (completed March 1994) using the Combined Arms and Support Task Force Evaluation Model (CASTFOREM), a constructive combat development model, and several high resolution scenarios to evaluate M1A1/M1A2 cost and operational effectiveness.[2]

The Warrior Focus AWE, which was one of the integral components of the Joint Venture Task Force initiative, was designed to explore the contributions of improved digitization and own-the-night systems and doctrine on the modern dismounted battlefield. The intent of this AWE was to use a series of live experiments, in concert with the MTM (or MEM, i.e., Model-Experiment-Model) methodology, to evaluate the digitization and own-the-night issues from the squad to battalion organizational levels. It was intended that the insights gained from this effort would be used to evaluate the JRTC 96-02 rotation at Fort Polk, LA, the final AWE event. The Dismounted Battlespace Battle Lab (DBBL) was responsible for directing the effort while TRAC-WSMR provided the analysis support.

M1A2 IOTE MTM APPLICATION

Figure 3 presents an overview of the modeling and analysis methodology developed and implemented by TRAC-WSMR used to support the M1A2 IOTE.

DEVELOPMENT OF TEST SCENARIOS

The M1A2 IOTE OPORDs were developed by the Armor School. The OPORDs and specific Fort Hood terrain data for each scenario were integrated into CASTFOREM which in turn produced estimates of the force exchange ratios and battle outcomes that were expected to occur for the M1A1 and M1A2 forces. The evaluation of the results focused primarily on two issues for each of the test scenarios. First, the desired defensive positions and corresponding 'avenues of approach' of the attacking force were optimized with respect to line-of-sight (LOS) to ensure maximum visibility. Second, if there were not considerable differences in the CASTFOREM estimates of

operational effectiveness between the M1A1 and M1A2 for a particular scenario, an analysis was conducted to determine if the scenario could be made more robust. The results of the pre-test scenario evaluation were provided to the test scenario development team for their review and subsequent integration.

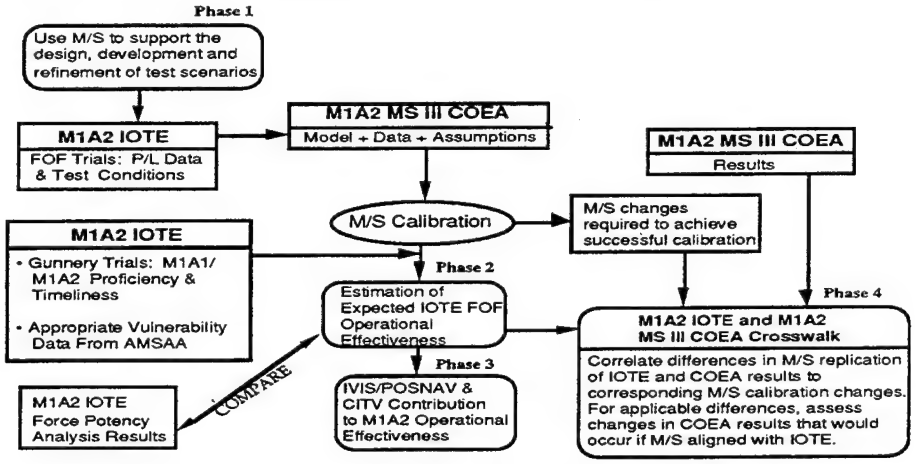


Figure 3. M1A2 IOTE MTM Methodology Overview

EXPECTED IOTE M1A1/M1A2 OPERATIONAL EFFECTIVENESS ESTIMATION

The operational effectiveness results from the M1A2 IOTE FOF trials were impacted by instrumentation problems which caused low firer-target shot pairing rates. Thus, systems would fire rounds but many would be assessed as misses when in reality, a hit occurred. Since the two opposing forces were not always killing (or being killed) at reasonable or consistent rates, the resulting operational effectiveness measures (e.g., number of shots, number of kills, number of losses, and loss exchange ratio) from the FOF trials were not as representative as they could have been. The poor pairing rates in the FOF trials also meant the classical M/S replication of the test operational effectiveness results could not be performed. Instead, it was decided to concentrate on that portion of the FOF trials that possibly could be replicated. Since a firer-target pairing only impacts the shot and kill assessment, it was assumed the remainder of the engagement process up to 'trigger-pull' should be repeatable using M/S.

Initially, the actual player location data and conditions (e.g., force composition/structure, weapon/sensor configurations, weather/atmospheric conditions, tactics and doctrine, etc.) from each of the FOF trials considered were integrated into the M1A2 MS III COEA version of CASTFOREM and corresponding data bases. An iterative calibration process was then performed to ensure that the various technical and tactical representations in the M/S aligned as closely as possible with what was actually occurred in the test. Termination of the M/S calibration process was achieved when the engagement profiles (i.e., the results of the engagement process up through 'trigger-pull') in the M/S replications and the test FOF trials appeared to be comparable. Cumulative time distributions and range/time scatter distributions of BLUE shots were used to define the engagement profiles. As an example, Figure 4 presents the shot range versus time scatter distributions that resulted in an FOF trial and the corresponding 21 M/S replications for one of the M1A1 defensive battles.

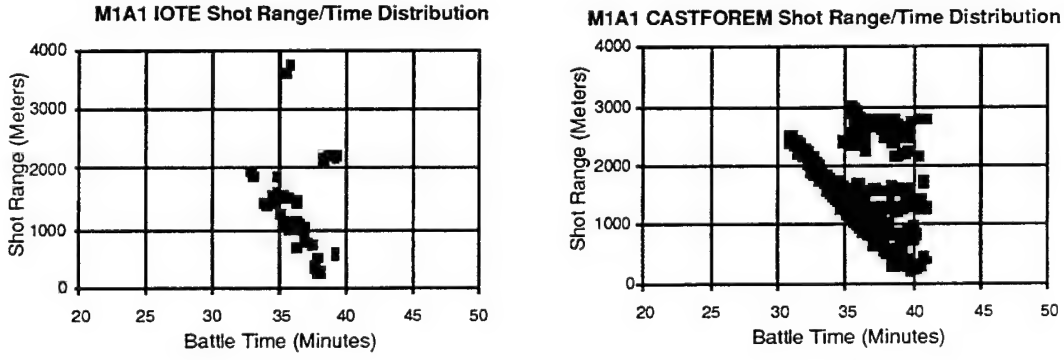


Figure 4. M1A1 Shot Range/Time Scatter Comparison Example

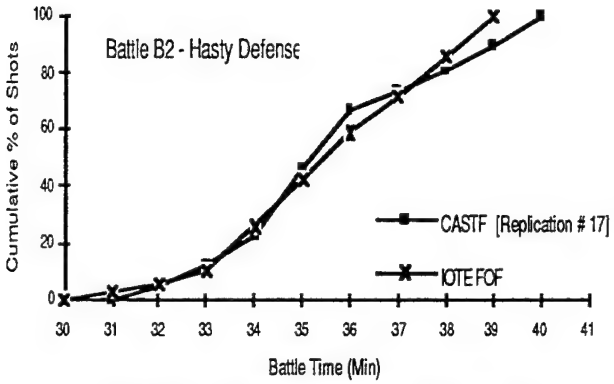


Figure 5a. Cumulative Shots Comparison

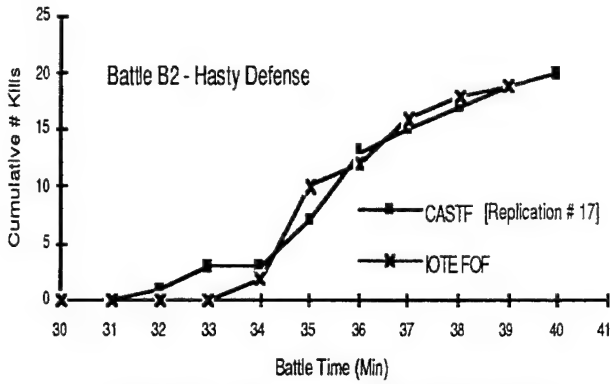


Figure 5b. Cumulative Kills Comparison

Figure 5a presents the cumulative percentage of all M1A1 shots over time resulting from a particular test FOF trial and a single M/S replication (for which the BLUE kill results are the closest to what occurred in the FOF trial) for the same M1A1 battle examined in Figure 4. Similarly, Figure 5b presents the cumulative number of M1A1 kills corresponding to the shots in Figure 5a. Although considerably more shots were fired in the test FOF trial than in the CASTFOREM replication (i.e., 143 versus 54), the relative BLUE force responsiveness (with respect to cumulative shots and kills over time) between the two is comparable. Note that the differences in kill levels between the test FOF trial and the M/S replication for a particular time interval were due to the differences in the outcomes of the stochastic draws for the P(Kill/Shot) events. Based on the engagement profiles and the shot and kill results examined for this M1A1 battle, it appears the engagement process simulated in the M/S replications was reasonably close to what occurred in the test FOF trial.

Because a reasonably close alignment between the M/S replication and test FOF trial engagement profiles had occurred, the M1A1 and M1A2 performance after 'trigger-pull' was evaluated. M1A1/M1A2 tank proficiency measures, i.e., P(Hit/Shot) and timeliness or interfiring times) were selected as the performance measures for comparison due to the availability of data. Since the results after 'trigger-pull' were questionable in the FOF trials for P(Hit/Shot), the tank proficiency results from the IOTE crew gunnery trials were compared with corresponding operational results in the M/S replications. The justification for this comparison assumes that the tankers should perform at the same proficiency levels in the gunnery trials as in the FOF trials. Interfiring times, however, could be compared between the FOF trials and the M/S replications because that data was not biased by the instrumentation shortfalls.

If there was a close similarity between the tank proficiency results from the gunnery trials and the M/S replications and if the appropriate lethality data were being used in the M/S, it was assumed that the M/S replication results should then provide a reasonable estimate of the 'upper bound' of the M1A1/M1A2 operational effectiveness that was expected to have occurred during the FOF trials.

For the most part, the M1A1 and M1A2 engagement profiles were reasonably close between the test FOF trials and the M/S replications. Some of the differences were caused by a considerable number of rounds being fired by the BLUE tanks at ranges exceeding 3000 meters in the test FOF trials and not in the M/S replications due to asystem performance data limitation. Also, some differences naturally occurred due to the effects of the BLUE force killing and being killed at a reasonably faster rate in the M/S replications than in the test FOF trials. In almost every case, the tank proficiency results from the M/S replications were somewhat greater than or close to what was demonstrated in the IOTE gunnery trials. Timeliness results showed the M1A1 and M1A2 interfiring times in the test FOF trials being somewhat shorter than in the M/S replications while the M/S replication and gunnery trial interfiring times were comparable. However, the M1A2 interfiring time advantage over the M1A1 was comparable between the test FOF trials and the M/S replications. The M/S replication operational effectiveness results aligned much closer with those from the Force Potency Analysis or FPA[3] (a procedure used by OPTEC to 'reconstruct' the test trials) than to the actual test FOF trial results. More specifically, the FPA operational effectiveness values were within the minimum and maximum CASTFOREM values for 86% of the measures examined. In the cases where the FPA and M/S replication results were not quantitatively close, the trends were similar.

IVIS/POSNAV AND CITV CONTRIBUTIONS TO M1A2 OPERATIONAL EFFECTIVENESS

For the third issue, the M/S test replication results were extended to assess the relative contribution of the Inter-Vehicular Information System/Position Navigation System (IVIS/POSNAV) and the Commander's Independent Thermal Viewer (CITV) subsystems to overall M1A2 operational effectiveness. The IVIS and POSNAV

subsystems are intended to enhance M1A2 performance with respect to maneuverability, situational awareness, and command and control (C2). The CITV subsystem, which allows the M1A2 to operate in a 'Hunter-Killer' mode, can enhance the M1A2 fightability characteristics (i.e., detect, engage, and kill at a faster rate) over those of the M1A1. Before the IVIS/POSNAV and CITV impacts on M1A2 operational effectiveness in the M1A2 FOF trials could be quantified, several assumptions were necessary. First, the IVIS/POSNAV representation in CASTFOREM allows the M1A2s to call in fire support faster and more accurately than do the M1A1s. However, other maneuver, situational awareness, and C2 aspects of IVIS/POSNAV were not explicitly represented in CASTFOREM. CASTFOREM does explicitly represent all 'Hunter-Killer' aspects of the CITV. It must also be assumed that part of the IVIS/POSNAV impact can be implicitly represented in CASTFOREM by the M1A2 maneuvering 'smarter' than the M1A1. This technique was successfully used in the M1A2 MS III COEA. Second, it must be assumed that the M1A1 and M1A2 M/S replications for a particular battle provided a reasonable estimate of the upper bound of what operational effectiveness was expected to occur (within the context of any explainable differences). Third, if the after-action report from an M1A2 test FOF trial (which was further substantiated by IVIS/POSNAV utilization by soldiers in the test) established that IVIS/POSNAV was used appropriately during that trial, it was assumed that the resulting M1A2 force maneuver data (which was used precisely by CASTFOREM) should reflect that usage of IVIS/POSNAV with respect to maneuver and situational awareness.

It was assumed that if the criteria discussed above constituted a valid premise, the following methodology (as summarized in Figure 6) could then be used to provide an estimate of the IVIS/POSNAV and CITV contributions to M1A2 operational effectiveness. First, the M1A1 M/S replication was rerun giving CITV capability to the M1A1 (designated by 'CITV + M1A1'). The CITV + M1A1 M/S results were compared to the M1A2 M/S replication results and the differences reflected the relative impact of IVIS/POSNAV on M1A2 operational effectiveness. Likewise, the CITV + M1A1 M/S results were compared to the M1A1 M/S replication results and the differences measured the relative contribution of the CITV to M1A2 operational effectiveness independent of the IVIS/POSNAV impact.

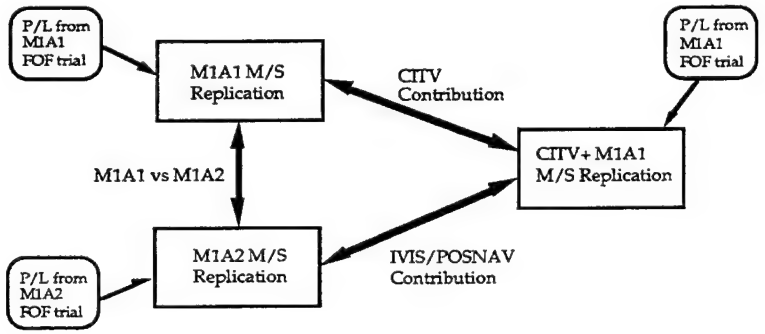


Figure 6. Estimation of IVIS/POSNAV and CITV Contribution

As an example, Figures 7a and 7b present the assessment of the impact of IVIS/POSNAV and CITV on M1A2 operational effectiveness for a BLUE defensive mission. The after action report for this M1A2 trial reported a moderate use of IVIS/POSNAV. As seen in Figure 5a, the results showed a considerable improvement in LER with a 80% increase in shots, a 9% increase in kills, and a 48% decrease in losses for the IVIS/POSNAV equipped force. However, in Figure 5b, the CITV equipped M1A1 force showed only a very slight increase in operational effectiveness over the M1A1 force.

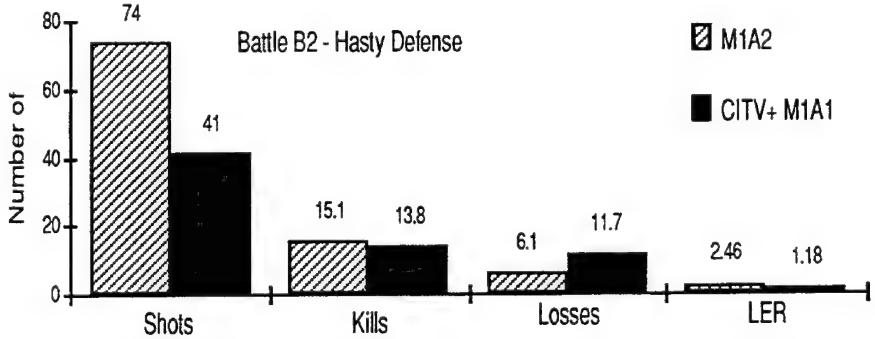


Figure 7a. IVIS/POSNAV Contribution To M1A2 Operational Effectiveness

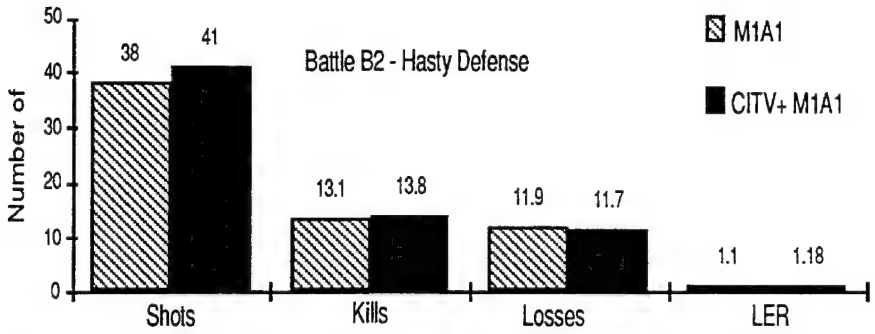


Figure 7b. CITV Contribution To M1A2 Operational Effectiveness

In general, the results revealed the IVIS/POSNAV contribution to M1A2 operational effectiveness to be considerable in each of the examined battles. The CITV results however showed minimal increases in M1A1 effectiveness when the M1A1 had the 'Hunter-Killer' capability. This implied the better maneuver and positioning demonstrated by the M1A2s (assumed to be due to IVIS/POSNAV) was the primary cause of the increase in operational effectiveness over the M1A1s. Note that the case with the CITV removed from the M1A2 (i.e., the M1A2 with only IVIS/POSNAV) was not evaluated and compared to the 'full-up' M1A2 case to assess the CITV contribution in an IVIS/POSNAV configuration. Since the IVIS/POSNAV subsystems provided the means for the Abrams tank to move 'smarter' (and thus be more effective and survivable), it was conjectured that there should be more targets for the M1A2 to engage in a typical battle as compared to that for the M1A1. If this indeed is the case, the CITV contribution to operational effectiveness may be somewhat greater when configured with the IVIS/POSNAV on the M1A2 than with no IVIS/POSNAV on the M1A1.

M1A2 IOTE AND M1A2 MS III COEA CROSSWALK

Linking or crosswalking operational tests and the corresponding COEAs is critical if the decision makers are to have consistent, credible, and meaningful information upon which to base their materiel acquisition decisions. However, linkage between these two diverse entities is not a trivial endeavor. Field test scenarios are normally quite restrictive with small forces (i.e., PLT to CO size on the BLUE side) due to limitations such as test range constraints, safety considerations, and cost. In contrast, COEA scenarios are very robust (normally at the BN/BDE level or higher) with a full combined arms 'flavor.' Because of these vast differences, it is virtually impossible to directly compare the results from an operational test and those from the corresponding COEA. The primary intent of the proposed linkage approach was to determine if the battlefield reality demonstrated in the field test was supported by and consistent with that demonstrated in the M/S. This was achieved by analyzing those changes to the COEA M/S required during calibration. If there were no apparent differences (i.e., no major changes to the COEA M/S, data, or assumptions were required), it could be conjectured (within the limited scope of the analysis) that the COEA and IOTE results were consistent and supportive of one another. If there were differences, the corresponding simulation changes were noted and analyzed to assess the subsequent impact on the actual COEA results if those changes were indeed applied to the COEA M/S and high-resolution scenarios.

After the P/L data and conditions for each test FOF trial were integrated into the COEA version of CASTFOREM (described in Issue 2), several inconsistencies in operational effectiveness were noted between the test FOF trials and corresponding M/S replications. In an iterative fashion, the M/S was calibrated by examining various aspects of the engagement process to determine the causes of those inconsistencies (summarized in Table 1).

The target acquisition process in CASTFOREM allows a weapon system's sensor(s) to scan its battlefield area of responsibility (i.e., field-of-regard or FOR) in a systematic fashion to find targets to engage. Each sensor scans its FOR, a field-of-view (FOV) at a time. The FOR may be defined to include areas where there are few, if any, targets (i.e., areas where the probability of a target being present is very low). In actual combat, tankers will follow a similar type of procedure to search for targets. However, the mental and optical processes used by the tanker result in a natural minimization of the time spent scanning areas where targets are not likely to be. To replicate this phenomenon in the M/S replications, the direction and size of each weapon's FOR in CASTFOREM was further optimized to realistically minimize the time expended searching in low target potential areas. M/S replication engagement results then showed a comparability with what had occurred in the test FOF trials. Note that the search process had been similarly optimized in CASTFOREM for the high resolution scenarios used in the COEA.

Normally in high resolution CASTFOREM scenarios, the defending force enhances its survivability by developing and utilizing some form of cover. If the situation allows enough time for a deliberate defense, the defensive vehicles will be dug in or in hull-defilade. If the time for the defending force to get into position is much

shorter (i.e., a hasty defense), the M/S user must decide what the appropriate level of coverage should be, i.e., half hull-defilade or fully exposed. There was some reference in the after-action reports that a portion of the M1A1/M1A2s performed berm drills (i.e., the tanks would back down a hill to hide and then 'pop back up' when they wished to engage) when the terrain allowed it. To represent the berm drill phenomenon in CASTFOREM, the BLUE defenders were assumed to be in a hull defilade posture once they reached their defensive positions. When BLUE was moving to or out of their defensive positions, however, they were placed in a fully exposed posture. Note that not all BLUE tanks in the defensive test FOF trials performed berm drills. Thus, placing the entire BLUE force in hull defilade in the M/S replications meant the defending M1A1s and M1A2s may have survived somewhat better than they did in the test FOF trials.

Table 1. M1A2 IOT&E and M1A2 MS III COEA Crosswalk Summary

Identified Difference Between Initial COEA M/S Replication & Actual IOT&E Results	Cause(s) of Differences	M/S Changes to Align With IOT&E [Calibration]	Impact on Consistency Between IOT&E and COEA
Lower firing level in M/S than in IOT.	M/S tankers may be spending too much time scanning low target potential FOVs.	Optimize FOR direction and size.	Engagement rates in IOT were somewhat greater than in initial M/S. Optimizing FORs in M/S will result in more realistic detection process as demonstrated in IOT and in COEA.
Inconsistent BLUE defending force postures.	In IOT, berm drills done as a survivability tactic. In M/S, BLUE defenders are in hull defilade. Defending forces in COEA normally in some level of defilade.	Blue defending forces assumed to be in hull defilade to reflect impact of berm drills.	Level of consistency of BLUE defending force survivability between IOT trials and M/S depended on to what extent the BLUE defenders performed berm drills.
BLUE tanks fire conventional rounds at ranges greater than 3000m in IOT but not in COEA.	In test, tankers can shoot at any range. In M/S, tank performance data is not defined past 3000m for conventional rounds.	Leave max range of 3000m in M/S.	Few, if any, shots fired in IOT at ranges greater than 3000m resulted in a kill. Thus the 3000m round limit should not impact the consistency between the IOT and COEA.
BLUE tank force engaged deeper and more effectively in COEA than in IOT.	In COEA, STAFF round was in M1A1/M1A2 basic loads but not in IOT.	STAFF round not played in IOT or in M/S replication of IOT.	If STAFF had been played in the IOT, the battle dynamics would have aligned more with that resulting in the COEA.
Tendency for M1A2 engagement capabilities to be overstated in M/S.	The TC cannot perform 'Hunter-Killer' 100% of the time as other functions must also be performed.	Set CITV utilization by TC at 80% which is what was used in COEA.	80% level of CITV utilization was consistent between M/S replication of IOT and COEA.
M1A1/M1A2 proficiency in M/S greater than or equal to that demonstrated in gunnery trials in most cases.	Accuracy data used in M/S replications may not reflect same tank proficiency as demonstrated in FOF trials	Use proficiency data as currently defined in M/S.	COEA and M/S replication M1A1 force effectiveness results may be slightly greater than what would have occurred in the FOF trials.
Interfiring times were somewhat shorter in FOF trials than in the M/S replications.	Tankers in FOF trials may have been firing at faster than expected rates due to the low pairing rates.	Since the interfiring times in the M/S aligned with the gunnery times, the M/S was not altered.	M/S interfiring results are consistent with COEA and gunnery results. IOT firing rates may have been more in line if perfect test conditions had resulted.

The M1A1/M1A2s in the test FOF trials often fired early long range shots in the 3000-4500 meters interval. These shots very rarely resulted in a hit and/or kill, partly because of the low pairing rates but also due to the very low hit and kill potential (i.e., PKSS) at those ranges. Performance data used in CASTFOREM to represent the M1A1/M1A2 hit and kill phenomena were not defined at ranges greater than 3000 meters. Thus, the 3000 meters threshold was used in the M/S replications and was consistent with what was used in the M1A2 MS III COEA. Note that if the 3000 meter range threshold had been extended in CASTFOREM for the COEA, there may have been an increase in shots fired but there would have been little, if any, difference in operational effectiveness because of the low PKSS at those ranges.

The M1A1 and M1A2 tanks in the COEA were equipped with STAFF rounds in addition to the standard conventional tank rounds. The STAFF round is a fire-and-forget, antitank munition that allows the BLUE tanks to engage targets more effectively and at significantly longer ranges than conventional tank munitions. The STAFF round was not part of the M1A1/M1A2 basic load in the IOTE trials. It is apparent then that operational effectiveness results, derived with and without the STAFF round, could be quite different. If STAFF had been played in the IOT, it is likely the battle dynamics would have aligned more with those in the COEA.

The M1A2 tank commander uses the CITV to hunt for new targets while the gunner is engaging (i.e., Hunter-Killer capability) which can dramatically impact the tank's target detection capability (especially as the number and dispersion of targets increases). The M1A2 tank commander, however, has to perform other functions (e.g., command and control, mission planning, etc.) which means he will not be using the CITV 100% of the time during the battle. To represent this phenomenon, the tank commander's CITV utilization level was set at 80% (derived and approved by the Armor School) which is consistent with what was used in the M1A2 MS III COEA.

The M1A1 and M1A2 operational hit proficiency demonstrated in the IOTE gunnery trials was slightly less than or very close to that occurring in the M/S replications. Since the M/S tank accuracy data were not modified to align with the gunnery results, the operational effectiveness in the M/S replications may have been slightly higher than what may have occurred in the test FOF trials under perfect test conditions. The interfiring times in the test FOF trials (especially for subsequent shots against the same target) are somewhat smaller than those in the M/S replications. Since the M/S replication interfiring times align with those in the IOTE gunnery trials, consistency between the test and the COEA prevails for this issue. This conclusion assumes that if the test conditions had been perfect, the interfiring times between the IOTE gunnery and FOF trials might have been more comparable.

The M/S changes, required to calibrate CASTFOREM to align with the test characteristics and conduct, collectively had a considerable impact on operational effectiveness results. However, in the context of the M/S modifications required during calibration, the operational effectiveness results from the initial M/S replications (which represent the COEA conditions and assumptions) and those using the calibrated M/S (which represent the IOTE conditions and assumptions) appear to be consistent and reasonable.

WARRIOR FOCUS AWE MEM APPLICATION

The general approach to the analytical support for the Warrior Focus AWE was to employ a series of integrated MEM phases to gain insights concerning the impact of enhanced digital and own-the-night systems on dismounted force effectiveness. Figure 8 presents an overview of how live experiments and constructive simulations were used as an analytical vehicle on which the AWE hypothesis and supporting global issues were evaluated.

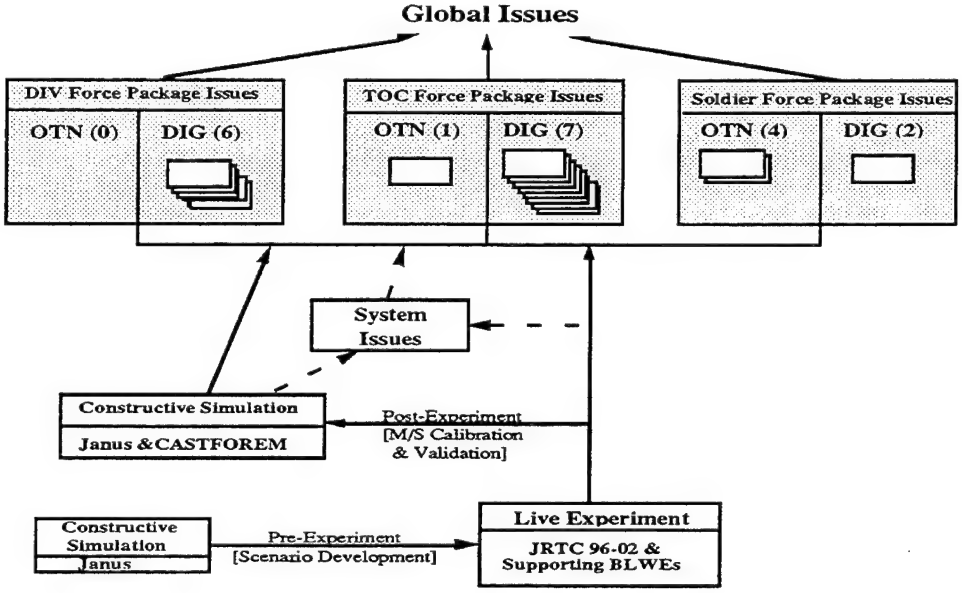


Figure 8. AWE Analysis Support Overview

Overarching global issues were addressed by evaluating six sets of issues that were categorized with respect to force level (DIV, TOC, and soldier). Each force level issue in a particular set applied to one or more systems that

were grouped by tactical function. Constructive simulations, i.e., Janus and CASTFOREM, were used as the primary modeling tools. Live exercises included the JRTC 96-02 rotation, supporting Battle Lab Warfighting Experiments (BLWE), and technical testing. The resulting data obtained during the conduct of live exercises were used in two ways. First, the data obtained from these events was processed to calculate MOP/MOE that were used to evaluate the appropriate AWE issues at the force level. The results also allowed SMEs to develop insights into the synergisms resulting from the employment of the enhanced systems and the corresponding TTP used in the simulations. These insights were then used to refine the TTP as required. This process is illustrated in an example of one of the AWE dendritics in Figure 9. Second, this data provided information for calibration/validation of the constructive simulations which ensured the application of the constructive simulations was credible and consistent with the field experiment results. The validated constructive modeling was used to address both critical system and force level issues that could not be directly addressed by the live experiments.

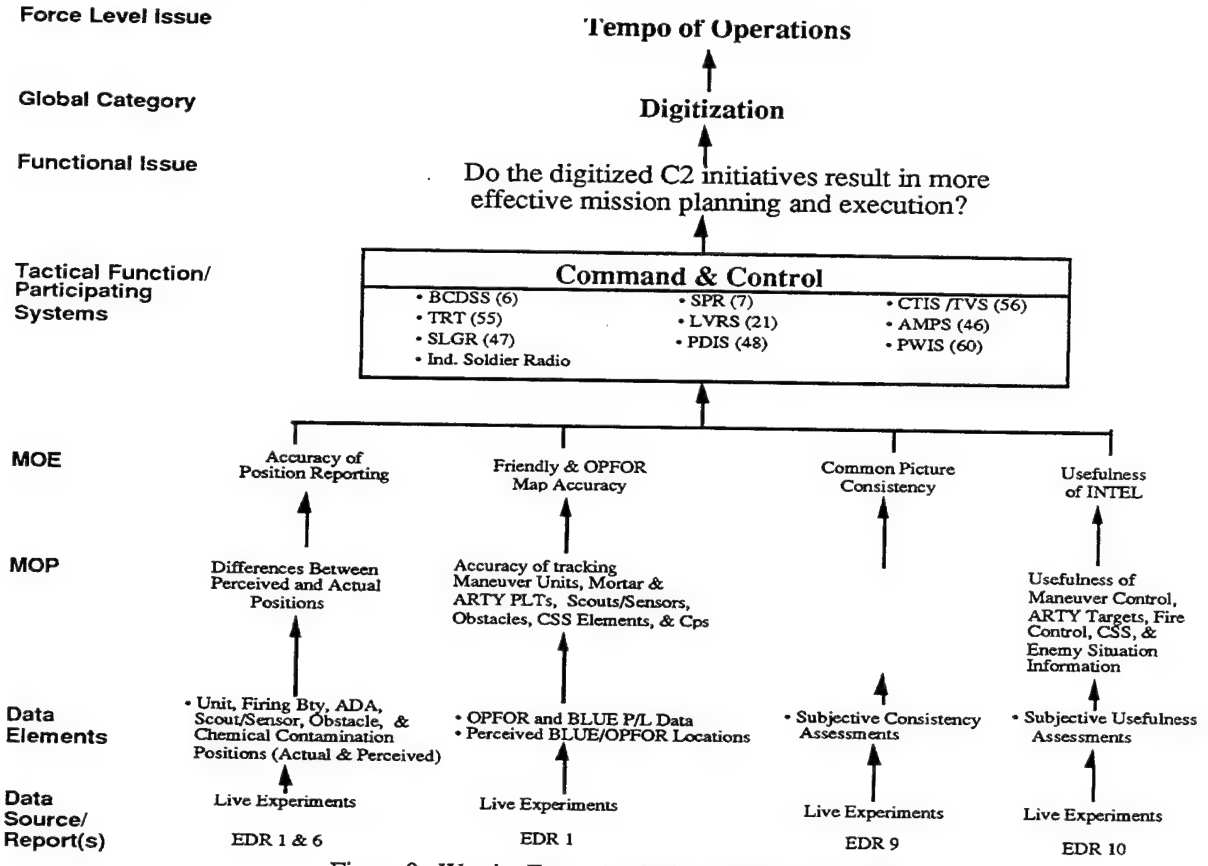


Figure 9. Warrior Focus Analysis Dendritic Example

The iterative MEM process was composed of a set of supporting BLWEs that examined AWE issues and objectives beginning at the soldier/section/ squad level up to battalion/TF level. The results from each level in turn provided the benchmarks for the subsequent echelon level analysis where additional systems were evaluated in concert with those already addressed at the previous level. Only those appropriate systems that could be realistically evaluated at each force level were included at that level. The iterative process concluded at the battalion/TF level resulting in the final validated constructive simulation required to support the AWE objectives and issues. Constructive simulations were used before each of the live experiments to develop and refine experiment scenarios. After each live experiment, constructive simulations were used to extrapolate and evaluate experimental results to those scenarios not addressed in the live experiments. and provide insights for the next iteration. Technical tests and additional experiments were used when possible.

It was initially planned that the appropriate exercise specifications that were to be used in the JRTC 96-02 (e.g., Fort Polk terrain, force structure and composition, TTP, etc.) would be integrated into the final validated constructive simulation to provide a prediction of what should have occurred in the JRTC 96-02 rotation under perfect conditions. After the rotation, the constructive simulations could have been calibrated to replicate what actually occurred in the field. The differences between the constructive simulations predictive results and the calibrated constructive simulation results, in terms of model modifications, then could be used to gain insights into

the results of the JRTC 96-02 exercise resulting in a better understanding of the representation of battlefield phenomena in constructive simulations.

Conduct of the Warrior Focus AWE was impacted by digital hardware and software problems due to the relatively early generation technologies that were fielded. These problems limited the scope of the live experiments which in turn limited the availability of experimental data thus constraining the usefulness of the modeling. However, the potential of these digital systems, even with the problems, became clear during the experiments as the enhanced digital processing and dissemination of information increased the situational awareness of the BLUE units. The experiments showed the own-the-night equipment to provide the BLUE dismounted forces a night range advantage and increases the ability to move faster at night resulting in a BLUE unit being able to seize the objective faster with fewer casualties.[4]

OBSERVATIONS AND CONCLUSIONS

Figure 10 summarizes several of the reciprocal and complementary relationships between operational testing and experimentation and M/S as demonstrated in the M1A2 IOTE and the Warrior Focus AWE. In these two efforts, M/S was used to assist the test scenario developers to design more robust and comprehensive test scenarios. M/S provided operational force effectiveness estimates that could be used to augment the test or experiment. The scope of the IOTE and AWE were extended to specifically quantify the contribution of particular systems to operational force effectiveness. Reciprocally, various aspects of the engagement process in the M/S were benchmarked against actual engagement results in live field tests and experiments. The M/S calibration procedure resulted in an intuitive understanding of the abstract representation of the many aspects of the engagement process in the M/S and how it relates to battlefield reality as demonstrated in the T/E. With such an understanding, predictive modeling results and T/E results could now be logically crosswalked based on any identified inconsistencies. While only two examples were specifically addressed here, it is hoped the results will provide insights into how the complementary aspects of operational testing and experimentation and constructive M/S can be used such that critical materiel acquisition issues concerning operational effectiveness can be better addressed in the future.

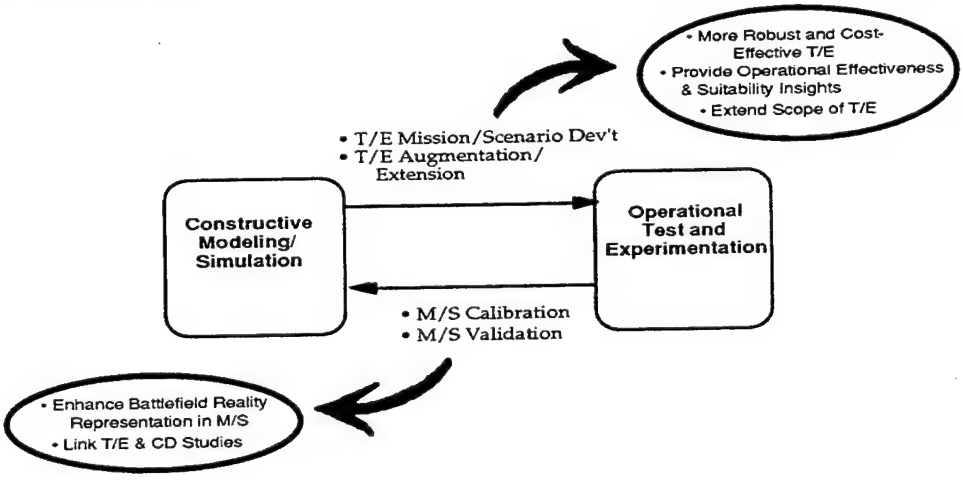


Figure 10. Complementary Relationships Between M/S and IOT

REFERENCES

1. Sanders, Dr. Patricia, "**Policy Guidance for the Application of Modeling and Simulation Support of Operational Test and Evaluation**", ITEA Journal ,Vol. X (1989), No. 3, OSD/DOT&E..

2. McCool, B. and Lyman, J., "**M1A2 IOTE Model-Test-Model: Post-Test Modeling/Analysis and COEA Linkage Results**", Draft Final Report, 1 June 1994, MTM Division, Fire Support Directorate, TRAC-WSMR.

3. Dubois, MAJ L. and Armendariz, V., "**Force Potency Analysis of RTCA Results**", 14 June 1994, OPTEC and Coleman Research Corp.

4. Eshelman, M., "Warrior Focus Hotwash", e-mail report of Warrior Focus Hotwash , 30 November 1995, TRADOC.

ON THE PERFORMANCE OF WEIBULL LIFE TESTS BASED ON EXPONENTIAL LIFE TESTING DESIGNS

Francisco J. Samaniego and Yun Sam Chong
University of California, Davis

ABSTRACT

It is common to plan a life test based on the assumption of exponentiality of observed lifetimes or lives between failures. Analysts are then able to calculate specifically how many items should be placed on test (or the number of observed failures it takes to terminate the test) and the maximum total time on test required to resolve the hypothesis test of interest. Once the test data is in hand, one has the opportunity to confirm the exponentiality assumption or to determine that an alternative modeling assumption is preferable. The question to be discussed here is "What if the data point to a nonexponential Weibull model?" Our purpose is to identify circumstances in which the available data permit testing the original hypotheses with better performance characteristics (that is, smaller error probabilities) than the test originally planned; a complementary analysis of situations leading to poorer performance is also given. Further, we will give an indication of the potential savings in the number of systems and the time on test that would accrue from having modelled the experiment correctly in the first place. Various approaches to Weibull life testing, with special attention to testing hypotheses concerning Weibull means, will be discussed. This paper is an expository presentation of issues and methods that are treated in detail in a manuscript with the same title that has been prepared as a commissioned paper for the National Academy of Sciences' Panel on Statistical Methods for Testing and Evaluating Defense Systems.

I. EXPONENTIAL LIFE TESTING

The widespread use of the exponential model in reliability and life testing studies is largely due to its great tractability. For many different experimental designs, it is possible to develop an exact analysis under exponential assumptions, and to determine, in advance, what resources are required to meet whatever bounds or requirements are set regarding confidence levels or error probabilities. Also, because many alternative lifetime distributions are "lighter-tailed" than the exponential, an exponential test will often be conservative, with both the producer's risk and the consumer's risk smaller than the nominal levels at which the test was planned. There are, however, a number of difficulties that arise when exponential analysis is used in non-exponential situations. First, in those cases in which the test is conservative, the opportunity to carry out a more efficient test, or to realize some savings in test resources, was foregone. Further, the nonrobustness of exponential life tests, especially when data is subject to some form of censoring, is well known. Of special relevance to us here is the robustness study of Zelen and Dannemiller (1961) which documents the failings of exponential life testing in Weibull environments. In short, a mis-applied exponential analysis can be dangerously misleading. It thus behooves the analyst to seek to discover when an exponential assumption is of dubious validity, and to execute an alternative analysis in such cases. In succeeding sections, we will focus on the particular alternative of Weibull life tests. In the remainder of this section we provide a brief review of the methodology of exponential life testing.

The problem of interest to us is the comparison of two means. We will assume that the null hypothesis that the true mean of the system under study is equal to M(0) is to be tested against the alternative hypothesis that the mean is M(1), where $M(1) < M(0)$. We will refer repeatedly to a document which describes in detail how such tests are carried out under the assumption of exponentiality: Department of Defense Handbook H108. As explained in that handbook, and elsewhere, one proceeds by fixing the operative levels of alpha (producer's risk) and beta (consumer's risk), after

Approved for public release; distribution is unlimited.

which one may identify the required number of observed failures r and the maximal amount of test time that would be required to resolve the test at the prescribed levels of alpha and beta. The test is then executed by rejecting the null hypothesis if the total time on test T at the time of the rth failure is less than some fixed threshold. The fact that the duration of the test, in real time, can be controlled, and made suitably small by placing n systems on test, where $n > r$, while resolving the test upon the rth failure, is an important side-benefit of the exponential assumption.

Since the mechanics of exponential life testing are quite well known, we discuss them only briefly here. The simplicity of exponential analysis of life testing data derives from the known and manageable distribution theory for the total-time-on-test statistic and from the fact that the characteristics of these tests depend on the hypothesized means M(0) and M(1) only through the so-called discrimination ratio $D = M(1)/M(0)$. Thus, the needed number of observed failures may be computed explicitly from the known value of the two error probabilities and this discrimination ratio. The required total test time may then be computed as a function of r, D, the error probabilities and the cutoff point of the appropriate chi-square distribution.

Implementation of an exponential life test design is facilitated by tabulations such as Table 2B-5 found in DoD Handbook H108. That table provides the values of the required sample size r and the ratio T/M(0) (and thus, indirectly, the required total time on test T itself) for selected values of D, alpha and beta; using it, one can set up and carry out exponential life tests with ease. As will be explained in the sequel, the table to which we've alluded is less than satisfactory for the kind of study we wish to do, namely, an examination of the performance of Weibull life tests when the true underlying life distribution is a nonexponential Weibull model. Handbook H-108's Table 2B-5 is simply too sparse to permit the identification of new values of r and T under these alternative circumstances. Some twenty pages of Samaniego and Chong (1995) are dedicated to the expansion of that table, and to computations based thereon. While these new tables will not be displayed here, we will comment on their construction and general characteristics, and will discuss the types of conclusions that one can draw from them. We first turn to a brief discussion of the Weibull distribution.

II. WEIBULL CONSIDERATIONS

The Weibull model is arguably the most popular parametric alternative to the exponential distribution in reliability applications. Like the gamma model, it contains the exponential distribution as a special case, so that the adoption of the Weibull assumption represents a broadening from the exponential model rather than a rejection of it. The parametrization we will use is denoted as W(A,B), where A is the "shape" parameter, that is, the exponent to which the lifetime t is raised in the exponential portion of the Weibull density, and B, raised to the power 1/A, is a scale parameter of the distribution. The mean M(A,B) and variance V(A,B) of a Weibull variable, in the parametrization above, is easily derived in closed form, each involving the scale parameter and certain gamma functions dependent on A. There are two elementary ways in which the Weibull and exponential models are related. First, the Weibull model W(1,B) is just the exponential model with scale parameter B. Secondly, and more generally, if X has the W(A,B) distribution, then X to the power A has the exponential distribution with scale parameter B. Further, the Weibull distribution has an increasing failure rate (IFR) when $A > 1$ and a decreasing failure rate (DFR) when $A < 1$. Another fact of special interest is that the coefficient of variation of the Weibull, that is, the ratio of its standard deviation to its mean, is independent of the parameter B. We will exploit this fact in the estimation of the shape parameter A in Section IV.

The statistical literature on modeling and inference based on the Weibull distribution is extensive. A keyword search of the Current Index to Statistics, v. 1-19, shows that there were 647 articles published in Statistics journals between 1975 and 1993 on Weibull-related topics. Surprisingly, very little of this work is directly applicable to the problem of interest here: tests for Weibull means in the general situation in which both parameters are unknown. Indeed, Lawless (1982) mentions that "life test plans under the Weibull model have not been thoroughly investigated...it is almost always

impossible to determine exact small-sample properties or to make effective comparisons of plans...further development of test plans under a Weibull model would be useful." It is our hope that the discussion of Weibull life testing is this paper will contribute to a better understanding of the possible advantages and risks these methods involve.

Before engaging in a Weibull analysis, one typically wishes to determine whether or not the Weibull model is reasonable in the application in which its use is contemplated. Without giving much detail, we mention some graphical techniques which have proven useful in explorations regarding the Weibullness of data.

Among the most revealing plot one can construct in such settings is the Total Time on Test (TTT) plot. This is essentially a plot of total time on test at any particular time t against the fraction of items in the sample that have failed by time t. To render TTT plots both manageable and comparable, a rescaled version of the total time on test statistic is generally used, resulting in a plot that fits inside the unit square. It is known that the theoretical TTT plot is linear for the exponential distribution, and is concave (convex) for distributions with increasing (decreasing) failure rate. We highly recommend TTT plots as a vehicle for recognizing nonexponentiality. While they don't point an analyst in the direction of a particular alternative model, a convex or concave plot is certainly suggestive that a nonexponential Weibull model should be examined as a possible alternative. A good reference on TTT plots is the paper by Barlow and Campo (1975).

A second graphical technique of interest, especially in the context of the present study, is that of the Weibull probability plot. These plots are based on the fact that, if S(x) is the survival or reliability function of the W(A,B) distribution, then $\ln\{-\ln S(x)\}$ in a linear function of the parameters lnB and A. It is thus possible to plot an empirical version of the relationship above to see if the data supports a hypothesized linearity. Such plots do give an immediate indication regarding the fit of the Weibull model. They provide, in addition, estimates of the Weibull parameters associated with the best fitting Weibull curve (fit according to the least squares criterion).

Through the use of graphical methods, or otherwise, assume that the analyst, after gathering data according to an exponential life test plan, determines that the data are more appropriately modelled with a nonexponential Weibull. It will then be necessary to proceed with an analysis appropriate for these broadened assumptions. The next two sections are dedicated to an examination of various ways of carrying out a Weibull life test.

III. WEIBULL LIFE TESTING - PART I

The classical theory of hypothesis testing yields its strongest results in problems in which the null and alternative hypotheses are simple, that is, specify the underlying probability model completely. In such problems, it is possible to construct optimal tests, that is, tests which minimize the consumer's risk beta among all tests with producer's risk less than or equal to some fixed level alpha. The problem of interest to us here is not of this type, and no "optimal" tests have been devised for solving the problem. When observable lifetimes are distributed according to W(A,B), an hypothesis that the mean is equal to some fixed constant K is equivalent to the statement that the parameter pair (A,B) lies in the subset of the first quadrant of the plane for which the Weibull mean satisfies the equation $M(A,B) = K$. This type of equation is a complex one, having no closed form solution. While the general problem is thus analytically challenging, there is a simpler problem for which an exact and optimal solution is available. We devote the present section to this simpler problem, with the goal of constructing a "gold standard" against which solutions to the more general problem can be compared.

Let us, then, suppose that a random sample $X(1),\ldots, X(r)$ of size r is drawn from what was originally thought to be an exponential distribution, and that the sample size r was determined from an exponential life test with fixed values of alpha and beta for testing the null hypothesis that the mean $M = M(0)$ against the alternative $M = M(1)$. Assume further that once the data was collected, the

W(A,B) model was adopted. Finally, let us make the simplifying assumption that the shape parameter A of the distribution is known. Then, by exponentiating to create "observations" of the form X raised to the power A, one can create a sample from an exponential distribution. The original null and alternative hypotheses may be rewritten as simple hypotheses in the parameter beta. In this new scenario, an optimal test exists: one should reject the null hypothesis M = M(0) if and only if the total-time-on-test statistic based on the new, exponentiated observations is smaller than a particular threshold. A very important change takes place in the process of making this transformation. The discrimination ratio in this latter problem changes to D raised to the power A. This change is significant in that the performance characteristics of exponential life tests depend very strongly on this parameter. The discrimination ratio is a measure of the distance between the null and alternative hypotheses. If that ratio is sharply reduced, the new testing problem can be resolved with much greater statistical power. Of particular interest to us is the fact that the same nominal values of alpha and beta can be achieved in the this new environment by a test based on a substantially smaller sample. It is important to note that an increase in the discrimination ratio results in just the opposite phenomenon. Thus, life testing in a Weibull environment is not necessarily advantageous to the tester. Fortunately, in many engineering applications of the Weibull distribution, the shape parameter A turns out to be larger than one, so that the opportunity for increased power and/or resource savings exists there.

A comprehensive analysis of the implications of a modeling shift from exponential to Weibull is not possible without new tables for exponential life testing which specify required resources for (essentially) a continuum of values of the discrimination ratio between zero and one. Because of its sparseness, Table 2B-5 in DoD Handbook H108 in not suited for our purpose. In Samaniego and Chong (1995), extensive tables are provided. Specifically, for four different alpha/beta pairs, tables are constructed showing sample size and total test time requirements for discrimination ratios in the range D = .01(.01).99. Further, for values of the (known) Weibull shape parameter in the range A = .1(.1)3, the values of four measures of performance of the test in the transformed Weibull problem are tabulated. These measures, and their definitions, are displayed below.

SSR = Sample Size Ratio = the ratio of the required sample size in the Weibull environment to the sample size in the original exponential life test, both computed to achieve fixed, predetermined error probabilities;

TTTR = Total Time on Test Ratio = the ratio of the maximum TTT required in the Weibull environment to the TTT in the original exponential life test, assuming fixed, predetermined error probabilities;

BR = Beta Ratio = the ratio of consumer's risks beta in the Weibull and exponential environments when the required sample size in the exponential environment is also used in the Weibull environment, with alpha fixed and equal in the two environments;

r/n = the ratio of the sample size in the exponential life test plan to the total sample size in a censored Weibull life test (terminated at the rth failure) which achieves the nominal error probabilities of the original plan.

The measures above are self explanatory, with the possible exception of TTTR. When one transforms Weibull data to exponential data via exponentiation, one can compute the new required sample size readily enough, given the new discrimination ratio, but the new required "TTT" is actually a sum of exponentiated Xs. We therefore need to determine how large the sum of actual failure times (that is, the sum of the Xs themselves) can be, given the value of the sum of exponentiated Xs. In Samaniego and Chong (1995), sharp upper and lower bounds are given for the actual TTT given the "TTT" of exponentiated failure times. When the Weibull shape parameter A is taken to be greater than one, the numerator of the measure TTTR is taken to be the upper bound of the actual TTT. Because of this, TTTR is a conservative measure, providing a figure that represents the guaranteed savings in test time had a Weibull life test been carried out to achieve the nominal error probabilities. The actual resource savings can, of course, be substantially greater than the bound this measure provides. When

the Weibull parameter A is taken to be smaller than one, the lower bound on the actual TTT is taken as the numerator of TTTR. This results in a bound which the actual TTT will exceed in the Weibull environment. Since the case $A < 1$ is somewhat rare in the types of applications we have in mind, we will not discuss that case further here.

An example of the use of the tables in Samaniego and Chong (1995) should be helpful at this point. Suppose one wishes to test the null hypothesis that the mean lifetime M of a system of interest is 1000 hours against the alternative that M is 500 hours (corresponding to a discrimination ratio of .5 if the exponential model were to be assumed). Suppose that the predetermined levels of alpha and beta are both .1. From the tables above, or from Table 2B-5 of Handbook H108, for that matter, one can determine that the resources required to resolve the test are $r = 15$ observed failures and a maximal TTT = 10,305 hours. Now, assume that it was determined in advance of the experiment that the Weibull W(2,B) model was applicable. It can then be determined that a life test in the Weibull environment, which is equivalent to an exponential life test with discrimination ratio $D = .25$, would require a sample size $r = 4$ observed failures, and a "TTT" of exponentiated Xs equal to 2,220,529.4 squared hours. From these facts, we find that

SSR = .267 and TTTR = .289.

It is thus apparent that substantial savings in both sample size and total test tine can accrue when the analyst is able to recognize a Weibull environment in advance.

Some commentary on the measure r/n is in order. As is well known, type II (or "order statistic") censoring has no affect on the analysis of exponentially distributed life test data. It's only impact on the test is the very welcome contribution it makes to the time it takes to complete the test. Unfortunately, the affect of censoring is different, and not as positive, in a Weibull environment. Indeed, when the censoring fraction is too small, the performance characteristics of a Weibull life test can suffer significantly. It is thus of interest to identify the amount of censoring that can be done in a Weibull life test that corresponds to a TTT requirement that is equivalent to that in the original exponential test plan. The ratio r/n, with r the required number of observed failures in the transformed Weibull problem, identifies the total sample size n which accomplishes this equivalence. In the numerical example above, we find from our tables that r/n = .084. From this, we deduce that a test plan which places 48 systems on test and resolves the test upon the 4th failure would have a total test time no larger than 10,305 hours, the test time associated with the exponential test plan based on 15 observed failures.

We close this section with some brief commentary on the major characteristics of the tables mentioned above. In general, these tables confirm the fact that there are potential resource savings available when one recognizes an IFR Weibull environment and carries out a Weibull life test instead of an exponential one. It will be clear from these tables that the most substantial savings in TTT are made in situations in which both the discrimination ratio and the Weibull shape parameter are high. It must be noted, however, that when the discrimination ratio is high (say, larger that .7), the costs associated with life tests are often exorbitant; thus, even though the resource savings afforded by a Weibull life test are substantial, the cost of the alternative analysis is still likely to be prohibitive. It appears that the kinds of problems in which recognizing a Weibull environment and performing a Weibull life test will be both feasible and economically viable will be those in which $.3 < D < .7$ and $A > 1.5$.

IV. WEIBULL LIFE TESTING - PART II

In the preceding section, we focused on the performance of Weibull life tests under the simplifying assumption that the Weibull shape parameter A was known. The assumption is not totally whimsical, since engineering experience with a particular type of application might make such an assumption quite reasonable. After all, the exponential assumption is nothing more than the

assumption that the Weibull shape parameter is known to be equal to one. It is, however, clearly necessary to move beyond this first step, and to engage seriously the question of how to execute a Weibull analysis in the general two-parameter problem. In this section, we will examine three specific possibilities in this regard.

The form of the optimal procedures in Section III for testing hypotheses concerning Weibull means immediately suggests a possible approach to the general problem: estimate the parameter A from data, and carry out a Weibull test as in the preceding section, with the estimated A taken as the known value of A. The performance of the resulting test procedure is naturally dependent on the quality of the estimator used. Results by Neyman (1959) and by Gong and Samaniego (1981) suggest that this "plug-in" approach tends to work well when "nuisance" parameters are replaced by root-n consistent estimators. Of particular relevance in the present problem is Neyman's paper, wherein the theory of C(alpha) tests was developed. Neyman gave conditions under which such tests were locally asymptotically most powerful.

The essence of a C(alpha) test is the substitution of one or more unknown parameters by root-n consistent estimators, and the testing of hypotheses concerning a lower dimensional parameter. We will examine two tests based on such an approach. The first is based on estimating the Weibull shape parameter A by a root-n consistent estimator derived as a function of the sample coefficient of variation. (See Sinha and Kale (1979) for a table relating the coefficient of variation of the Weibull model to its shape parameter.) The second test is based on estimating A by the root-n consistent estimator obtained from the best-fitting Weibull distribution from Weibull probability plots. The asymptotic properties of estimators based on probability plots have been studied by Nair (1984).

The performance of a third procedure for testing between competing Weibull means has also been studied. This is the likelihood ratio test for the null hypothesis that the mean $M = M(0)$ against the one-sided alternative that $M < M(0)$. The numerical issues that arise in executing this test are discussed in detail by Samaniego and Chong (1995). Since the likelihood ratio statistic is expected to be large under departures from $M = M(0)$ in either of two directions, we executed this test by doubling the nominal tail probability and rejecting the null hypothesis only when the data is indicative of a mean value smaller than M(0).

Samaniego and Chong (1995) report on an extensive simulation in which the performance of each of the three tests above is compared to the performance of the best test possible, that is, the uniformly most powerful test when the shape parameter A is known. We report briefly on our findings. First, it is clear that when the underlying Weibull distribution is strongly DFR, that is when A is quite near zero, Weibull life testing is nearly hopeless. Even tests which exploit knowledge of the true value of A have low power at the alternative hypothesis. Since DFR Weibull models are of relatively little interest in most life testing situations, this deficiency of the tests examined is not particularly worrisome. Our primary interest is in the behavior of our three general tests, as compared to the "gold standard", when the true shape parameter A is larger than one.

The most surprising and encouraging aspect of our simulation study is the fact that, when A is an unknown value greater than one, the three procedures for testing means in the general two parameter problem each performs nearly as well as the best test when A is known. As an example of the surprising competitiveness of a two-parameter Weibull life test, consider testing $M(0) = 1000$ against $M(1) = 500$ at alpha = .1. Suppose fifteen systems are placed on test, as prescribed by an exponential life test plan with alpha = beta = .1. If the data happens to be governed by a Weibull distribution with shape parameter $A = 1.2$, and the fact that $A = 1.2$ is somehow revealed to the experimenter, the best test can executed by exponentiating, that is, raising each failure time to the 1.2 power, and applying an exponential analysis on the transformed data. Our simulations show that this procedure achieves approximate error probabilities alpha = .11 and beta = .02. Now, suppose A was in fact not known. How well would the analyst do using any of the general tests? The answers are displayed below.

cv-based test: alpha = .11, beta = .02,

Weibull plot-based test: alpha = .09, beta = .03,

lr test: alpha = .13, beta = .02.

The performance of all three procedures is clearly indistinguishable from that of the best test in this example. A general perusal of the error probabilities in the tables reporting our simulation shows that this example is not an isolated instance of this type of performance.

V. DISCUSSION

As a general conclusion, it seems reasonable to state that our study strongly supports that claim that Weibull life testing can be both analytically feasible and economically effective. Our interest in the general question of efficient alternatives to exponential life testing was rekindled by the recent workshop on Defense Testing co-sponsored be the Department of Defense and the National Academy of Sciences' Committee on National Statistics. (See Rolph and Steffey (1994).) The optimistic conclusion we have reached with regard to Weibull life testing must be tempered with some cautionary words. First, it should be noted that our study is predicated on the tacit assumption that the analyst has determined that a Weibull model is appropriate in a particular application. When this is the case, and it is also determined that the applicable shape parameter is larger than one by at least some specific positive amount, then one may indeed take advantage of the resource savings available from using a Weibull test instead of an exponential test. The inclination to employ an IFR Weibull model without serious justification must be avoided. The allure of potential resource savings must not obscure the dangers and costs of model mis-specification.

The studies reported in Section IV above are based on complete rather than censored samples. When this work was presented at the First Army Conference of Applied Statistics in October, 1995, the influence of censoring on these results was not yet known. Since that time, some simulations based on type-II censored data have been completed. Samaniego and Chong (1995) discuss this additional Monte Carlo study. In brief, we may summarize our findings by stating that excessive censoring affects the power of Weibull life test adversely, but that, under moderate censoring, the performance of Weibull life tests compare quite favorably to the aforementioned "gold standard".

REFERENCES

1. Zelen, M. and Dannemiller, M. "The Robustness of Life Testing Procedures Derived from the Exponential Distribution." Technometrics 3, 29-49, 1961.

2. Handbook H108: Sampling Procedures and Tables for Life and Reliability Testing (Based on the Exponential Distribution). New York: John Wiley and Sons, 1960.

3. Samaniego, F. J. and Chong, Y. S. "On the Performance of Weibull Life Tests Based on Exponential Life Testing Designs." Technical Report No. 316, Division of Statistics, University of California, Davis, 1995.

4. Lawless, J. F. Statistical Models and Methods for Lifetime Data. New York: John Wiley and Sons, 1982.

5. Barlow, R. and Campo, R. "Total Time on Test Processes and Applications to Failure Data Analysis." In Barlow R.E., Fussell, R. and Singpurwalla, N. D. (editors), Reliability and Fault Tree Analysis. Philadelphia: SIAM, 451-81, 1975.

6. Neyman, J. "Optimal Asymptotic Tests of Composite Statistical Hypotheses." In Grenander, U. (editor) Probability and Statistics, The Harold Cramer Volume. New York: John Wiley and Sons, 1959.

7. Gong, G. and Samaniego, F. J. "Pseudo Maximum Likelihood Estimation: Theory and Applications." Annals of Statistics, 9, 861-69, 1981.

8. Sinha, S. K. and Kale, B. K. Reliability and Life Testing. New Dehli: Wiley Eastern Limited, 1979.

9. Nair, V. "On the Behavior of Some Estimators from Probability Plots." Journal of the American Statistical Association, 79, 823-31, 1984.

10. Rolph, J. and Steffey, D. (editors) Statistical Issues in Defense Analysis and Testing: Summary of a Workshop. Washington D. C.: National Academy Press

PENETRATION AND DEFLATION ALGORITHMS FOR TIRE VULNERABILITY

Ricky L. Grote, Linda L. C. Moss, and Edwin O. Davisson
U.S. Army Research Laboratory
Aberdeen Proving Ground, MD 21005-5068

ABSTRACT

An improved methodology was developed for the U.S. Army Research Laboratory's Ballistic Vulnerability/Lethality Division to conduct vulnerability analyses of wheeled vehicles subjected to attack from small steel and tungsten fragments less than 0.65 g (10 gr). Sequential statistical methods were used to collect and compute the velocity at which 50% of the fragments are expected to perforate the target, V_{50}. Also the traditional THOR penetration equations were improved upon for tires. Methodology was developed to predict tire deflation with and without an on-board Central Tire Inflation System and to determine the level of damage required to render a tire or tires nonfunctional within given time limits.

1. INTRODUCTION

The U.S. Army Research Laboratory's Ballistic Vulnerability/Lethality Division (BVLD) reviewed the current methodologies used to conduct vulnerability analyses of wheeled vehicles subjected to attack from steel and tungsten fragments. This internal audit revealed that there was a lack of experimental data that would defend the use of existing tire probability of kill functions (now called probability of component dysfunction, P_{cd}) and support the use of current penetration equations when "small" fragments are considered. Additionally, the capability of an on-board Central Tire Inflation System (CTIS) had never been considered explicitly in determining the level of damage required to render a tire or tires nonfunctional within given time limits.

Consequently, an experimental program plan was developed with the goal of producing sufficient data to substantiate the current tire vulnerability methodology or to allow for development of new methodology, if necessary. The tire vulnerability modeling process proceeded by answering the following questions in order:

1. Do small fragments have the ability to perforate tires?
2. What are the effects of multiple perforations in a single tire?
3. If tires are perforated, what is the resulting deflation rate?
4. If the target has a CTIS, how does CTIS performance influence deflation rate?
5. What are the effects of perforations in multiple tires on a vehicle with a CTIS?

The experimental plan was developed to directly address these five questions. The ability to model fragment perforation of tires was to be addressed by adding small steel and tungsten fragment data to an existing data set for the development of an improved and more general penetration equation and by determining V_{50} ballistic limits for various tire cross sections. The remaining questions which pertain directly to the validity of the existing P_{cd} functions were to be resolved via experimental firings at pressurized tires mounted on a vehicle with a CTIS and theoretical development of governing equations. A Soviet BM-21 multiple rocket launcher (MRL) was selected as the target vehicle for several reasons: it was available, it contained a CTIS, and additional tires were available for testing.

2. TARGET VEHICLE

The BM-21 MRL consists of a launcher assembly mounted on a URAL-375D chassis. The launcher assembly contains 40 firing tubes for high-explosive-fragmenting munitions. Most of the discussion will focus on the URAL-375D, since the tires and the CTIS are part of this truck. The URAL-375D is a 4.5-ton, three-axle 6×6 cross country vehicle that may be used on surfaced roads, earthen roads, and on roadless terrain. The URAL-375D also incorporates the use of adjustable inflation tires and a CTIS for increased mobility. The service manual (USSR, undated) for this vehicle provides guidance for tire pressure settings and driving speeds for various road surface conditions.

2.1 TIRE CHARACTERISTICS

The tires mounted on the BM-21 are 14.00-20 adjustable inflation tires, model 01-25. These tires consist of 10 plies, are tubed, and have the typical cross-country tread design. The tire cross-sectional thickness varies both in the tread and sidewall areas. The sidewall ranges from about 1.52 cm (0.6 in) to 2.54 cm (1.0 in) and the treads vary from about 5 cm (2 in) to 6.6 cm (2.6 in) with the area between the tread lugs around 2.54 cm (1.0 in).

2.2 CTIS CHARACTERISTICS

Soviet-designed tire-inflation systems were first included on pure transport vehicles in 1958. Two basic types of CTIS's were implemented in vehicle designs. The oldest type, introduced on the ZIL-157 and resident on the BM-21, is more complex and offers more flexibility to the operator (Warner 1987). This system contains control valves that can isolate any individual tire or set of tires. The pressure in the tires with open control valves can then be adjusted either up or down through the use of a three-way slide valve that has positions for inflate, deflate, and neutral (Warner 1987). The newer design does not have the cab-mounted valves for each tire. In this case, air supplied through use of the three-way valve is fed into a manifold and onto each axle of the vehicle from a common supply. The GAZ-66, ZIL-131, URAL-1320, and KrAZ-25TB are known to use this system.

3. V_{50} BALLISTIC LIMITS

3.1 V_{50} EXPERIMENTAL PLAN

The intent of the ballistic limit experiments was to answer, at least in part, the first question listed in the introduction: Do small fragments have the ability to perforate typical combat vehicle tires? V_{50} is that velocity at which 50% of the fragments are expected to perforate the target. The V_{50} ballistic limit data are very useful in that they provide a quick notion of what threat mass and velocity are required to defeat a target. It can also be used in some penetration equations to provide predictions of residual velocity. The experimental matrix included a range of small fragment sizes, 0.13 to 0.37 g (2 to 5.7 gr), and tire thicknesses of 1.52 to 6.6 cm (0.6 to 2.6 in).

A single tire was used for the V_{50} and penetration equation portions of this effort. The tire was cut into 16 wedge-shaped sidewall pieces and 8 tread sections. Each section was rigidly clamped to a test stand so that the rubber would be somewhat rigid.

The experimental setup for the V_{50} and penetration equation work consisted of a 5.56-mm gun, four velocity breakscreens, and a stand to hold the tire sections. A piece of photo paper taped to the front of the tire section was used for checking fragment orientation at impact.

The experimental procedure followed for the V_{50} work was the Up and Down Method, which is described in Darcom Pamphlet 706-103 (U.S. Army Materiel Development and Readiness Command 1983), AMC Pamphlet 706-111 (U.S. Army Materiel Command 1969), and JMEM Surface-to-Surface Manual JTCG/ME-61S1-3-4 (Joint Technical Coordinating Group for Munitions Effectiveness 1982). Basically, the velocity is increased or decreased incrementally depending on whether or not perforation is achieved. Experimentation stops when a specified number of firings have been conducted or when a zone of mixed results is achieved.

3.2 V_{50} ANALYSIS AND RESULTS

The DiDonato and Jarnagin procedure (McKaig and Thomas 1983) was implemented to obtain unique maximum likelihood estimates of the mean and the asymptotic standard deviation of the perforation distribution for various threat masses against target thicknesses. The V_{50}, also known as the ballistic limit, is the mean of the perforation distribution. Also determined was the standard deviation of V_{50}, which is a measure of the accuracy of V_{50}. Simply stated, if additional data sets were provided with the same mass against the same target thickness, the V_{50} calculated could vary from the one computed for the previous data set. The standard deviation of V_{50} indicates the amount of variability in the V_{50} estimate.

Unique maximum likelihood estimates are possible as long as two restrictions hold. The first restriction requires a zone of mixed results, in which the lowest velocity that perforated the target is smaller than the highest velocity that did not perforate. The second restriction requires that the average velocity for the perforated data is greater than the average velocity for the nonperforated data. When these restrictions did not hold, the standard deviation of the V_{50} calculation was not possible. The V_{50} and asymptotic standard deviation estimates were then obtained using a nonparametric method. This method used the three highest velocities that did not perforate with the three lowest velocities that did

perforate. There were 279 firings conducted to obtain the data required for the V_{50} calculations. Of these, 131 data points from perforating shots could also be used in support of the penetration equation development.

Comparing the perforation capabilities of steel and tungsten, given the same fragment mass against the same tire thickness, demonstrated that tungsten has a lower V_{50} than steel, more than 100 m/s lower (see Table 1). This is not surprising since tungsten has a higher density (17.7 g/cm^3) compared to steel (7.8 g/cm^3).

Table 1. Some V_{50} Comparisons

Mass (g) [gr]	Tire Thickness (cm)	V_{50} (m/s)			
		Steel		Tungsten	
		Mean	Std. Dev.	Mean	Std. Dev.
0.13 [2]	1.52 Sidewall	550	*	330	6.74
	2.54 Sidewall	704	27.54	471	6.00
	2.54 Between Lugs	N/A	N/A	434	2.20
0.26 [4]	1.52 Sidewall	399	14.16	293	*
	2.54 Sidewall	621	11.07	413	5.94
	2.54 Between Lugs	624	52.52	382	8.73

N/A = 0.13 g between lugs was not tested.

* Standard Deviation V_{50} estimate not available, since there was not a zone of mixed results.

4. PENETRATION EQUATIONS

4.1 PENETRATION EQUATIONS EXPERIMENTAL PLAN

Danish (1968, 1973) had already developed penetration equations from experimental data that included steel right circular cylinder (RCC) fragment simulators ranging in mass from 0.32 to 7.78 g (5 to 120 gr) fired against various tire thicknesses. The intent of the experimental design developed for this effort was to supplement the work of Danish with smaller RCC firings, from the V_{50} work, and with approximately 50 firings of real fragments so that the validity of the penetration equations could be extended to the smaller fragment regime. The mass of the real fragments ranged in size from 0.05 to 0.36 g (0.77 to 5.5 gr). They were fired against the sidewall, lugs, and between lugs. For the development of the penetration equation, the procedure required that all firings produced fragments that perforated the tire sections. Experiments were conducted with different striking velocities so that a range of overmatches was achieved for each fragment mass and target thickness combination.

4.2 PENETRATION EQUATIONS ANALYSIS AND RESULTS

In the 1960s, The Johns Hopkins University developed a set of empirical penetration equations based on steel fragments fired against various materials, including rubber. This effort, called Project THOR (The Johns Hopkins University 1961), predicts residual velocity and residual mass given the following independent variables: target thickness, average impact area of fragment, fragment striking mass, obliquity, and fragment striking velocity. Coefficients were computed for each of the different target materials. The forms of the equations are as follows:

$$V_r = V_s - 10^a (TA)^b M_s^c (\sec\theta)^d V_s^e \quad (1)$$

and

$$M_r = M_s - 10^f (TA)^g M_s^h (\sec\theta)^i V_s^j, \quad (2)$$

where V_r = residual velocity (fps); V_s = striking velocity (fps); T = thickness of target (in); A = average impact area (in^2); M_r = residual mass (gr); M_s = striking mass (gr); θ = obliquity angle (deg); a, b, c, d, e, f, g, h, i, j = empirically determined coefficients.

Danish (1968, 1973) realized that the use of rubber as a target material for tires was inappropriate, since tires have nylon threading in addition to rubber. Therefore, he used the THOR form to update coefficients for the penetration of steel fragments against tires.

Danish claimed that during his experimentation, fragment mass did not degrade when perforating tires. This claim was substantiated in the very early firings conducted as part of the ballistic limit work. Thus, only a residual velocity

penetration algorithm was deemed necessary. The experiments conducted by Danish included 641 fragments with the range of mass size from 0.32 to 7.78 g (5 to 120 gr). Both bias-ply and radial tires were included in the combined data set, since statistical analysis showed they were not significantly different.

Our goal was to augment Danish's data with smaller fragments and to account for fragment material differences. Data for the algorithm development included data from 131 firings that perforated the target in the ballistic limit experiment and 36 additional firings of real fragments along with 641 data from Danish's experiments. This provided a total of 808 data points for the development of a THOR-type penetration equation.

The THOR form (Equation 1) is nonlinear, and the coefficients could be determined using nonlinear least squares. However, the nonlinear least squares simply performs a fit, and it is difficult to test for the significance of the estimators. Except asymptotically, nonlinear regression does not provide the ideal statistical properties of unbiased and minimum variance estimators, as linear regression. Also, there are no exact statistical tests on model parameters for nonlinear regression (Myers 1990).

Although the THOR form is a nonlinear equation, it is intrinsically linear, since it can be transformed into a linear form:

$$\log (V_s - V_r) = a + b \log (TA) + c \log (M_s) + d \log (\sec \theta) + e \log (V_s). \quad (3)$$

Thus, the statistically significant variables and their corresponding coefficients can be estimated through the use of multiple linear regression. The original THOR project conducted by the Johns Hopkins University proceeded in this manner. Variables considered in the current anlaysis included the original THOR variables, separating the variables thickness and area, and three variables that might account for threat material differences. They are K = shape factor ($cm^2/g^{(2/3)}$), D = density of material (g/cm^3), and E = modulus of elasticity (megapascals). However, only one of these three variables is necessary to describe material differences. The correlation between density and modulus of elasticity is 1.00; therefore, we dropped elasticity, since density is an easier variable to obtain. The correlation between density and shape is –0.76 (in log scale). Either one (but not both) could be used in the model. The adjusted R^2 value (R^2_{adj}) using either variable is 0.783. R^2 is the ratio of the variation of the regression sum of squares for a given regression model to the variation of the total sum of squares for a given data set. The closer this ratio is to unity, the more efficient the model is at prediction. The adjustment to R^2 accounts for the degrees of freedom in the model, and thus, allows for proper comparisons among models with different numbers of independent variables. The variance inflation factors, eigenvalues, and conditioning index for shape or density with the other significant variables are well within the rule-of-thumb criteria (Myers 1990) for checks of ill-conditioning. These are measures of correlation between the regressor variable to enter the model and the variables already in the model when performing a stepwise regression procedure.

Shape was chosen over density since shape factor is more intuitive to a vulnerability analyst as an indication of a fragment's ability to penetrate. For example, a rod will perforate a target easier than a sphere of the same mass and velocity. Using shape also meant that a vulnerability analyst would only need typical arena data for fragmenting munitions and would not have to research material properties such as density.

4.2.1 Obliquity. The final form of the tire penetration equation does not include a term for target obliquity, since all firings were conducted at a 0° obliquity. It was felt that obliquity would not be a significant parameter for a "soft" target except for the increase of the target line-of-sight thickness. Danish (1968) had conducted four firings with 3.9-g (60 gr) steel fragment simulators at an obliquity of 60° and came to the same conclusion about the significance of obliquity.

To further investigate the effect of obliquity, sixteen additional firings that perforated the 1.52-cm sidewall target were conducted at a 45° obliquity. Three were steel and 13 were tungsten RCC fragment simulators; both types were 0.26 g (4 gr). Using the penetration equation, a check for consistency was conducted by changing the line-of-sight thickness (by multiplying thickness by $\sqrt{2}$) and using the new model to predict the residual velocity. The standardized residuals are all within ± 2.2. Generally, if the residuals are random and within ± 3.00, the model is in check.

4.2.2 Real versus Simulated Fragments. Both real and simulated tungsten fragments were included in the data set of 808 points, providing a good opportunity to determine whether there is a significance between the two for developing penetration equations. The total number of 112 tungsten fragments included 45 real and 67 simulated fragments. An indicator variable in the regression analysis revealed that there is no significant difference between real and simulated fragments.

4.2.3 Multiple Barriers. The penetration equations developed under the original THOR project were for perforation of a single target plate. Over the years, the THOR equations have been applied recursively to successive

plates even though the additional plates are beyond the limits of the data from which the equations were generated. Much of the error associated with this practice comes from the fact that fragments are deformed and change shape when they perforate metal plates. Thus, the shape of the fragments at successive plates would be changed from the original.

Bely et al. (1992) showed that, theoretically, implementing a recursive algorithm will not provide an unbiased estimate of the final residual velocity. Their analysis showed that the residual velocity underpredicted approximately 50 m/s for each metal plate perforated. However, tires are "soft" targets, and it was shown via experimentation that fragment mass did not change upon perforation; therefore, it is possible that successive barriers may be handled by the penetration equation in either a recursive fashion or by increasing the thickness term. In an attempt to gain insights into this issue, a small excursion was conducted. Eighteen firings were conducted at the sidewall of an intact tire. These 18 firings resulted in two 0.32-g (5 gr) and five 0.97-g (15 gr) steel fragment simulators that perforated when fired at the 1.52-cm area of the sidewall. The intent was to perforate both sidewalls of the tire and to record striking and residual velocities. The tire was not inflated so that a single tire could be used for this excursion.

The penetration equation was first implemented in a recursive manner to check agreement with the experimental data. The residual velocity computed from the equation after penetration of the first side of the tire was used as the striking velocity for the second tire barrier. The predicted versus observed plot for the final residual velocity revealed no bias in the prediction, and the data fit well around the perfect fit line. When the tire thickness was doubled as input into the penetration equation, the predicted residual velocity consistently overpredicted the actual value. Although this excursion was taken on a very small sample, it does allow us to see that there are no gross errors for recursively estimating a fragment through multiple barriers, when the target is "soft" relative to the fragment.

4.2.4 Coefficients and Goodness-of-Fit. The significant variables and their coefficients solved in the linear form were transformed back into the original form as

$$V_r = V_s - 10^{1.479}\ K^{0.411}\ M_s^{-0.191}\ T^{0.732}\ V_s^{0.350}. \qquad (4)$$

The fit of the equation to the experimental data is reasonable as given by an adjusted R^2 of 0.783. The standard error, also known as the square root of the residual mean square error (MSE), $\sigma_{\{\log\ (vs\ -\ vr\}}$, is 0.102. Both the R^2 and the standard error are comparable to Danish's original fits with values of $R^2 = 0.752$ and standard error = 0.096.

Figure 1 presents a predicted versus observed residual velocity plot, which shows that there is reasonable agreement between the Equation 4 and the observed data set. If the model were a perfect representation of the data, all points on the graph would fall on the solid, perfect fit line. It is obvious, by inspection, that the points do tend to cluster around the perfect fit line, with slight overpredictions when V_r is close to 0. (These overpredictions close to 0 can be corrected with the incorporation of V_{50} to the THOR penetration equations. This is addressed in the expanded ARL report [Grote et al. 1996].) One datum, with a standardized residual of -11.1, is encircled as an outlier. This was also an influential point and therefore was omitted in this final fit. Other points that appear to stray from the perfect fit line have smaller standardized residuals in the logarithmic scale. Omitting them did not substantially change the equation and thus were not highly influential. Therefore, they remained in the data set for the model development.

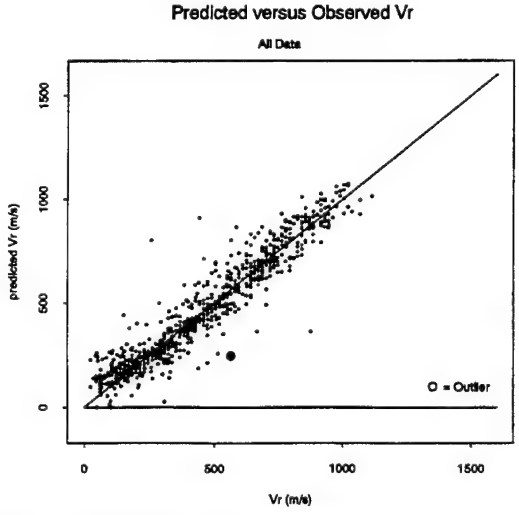


Figure 1. Predicted versus Observed Residual Velocity, All Data.

5. TIRE DEFLATION AND CTIS PERFORMANCE

5.1 DEFLATION AND CTIS EXPERIMENTAL PLAN

The objective of the tire deflation effort was to determine how quickly tires would become inoperable in the absence of a functional tire inflation system. A set firing matrix was not developed prior to initiation of this effort due to the uncertainty of what combination of number of fragment perforations and sizes of fragments would be required to deflate the tires within the time limits prescribed by traditional A-, B-, and C-kills. Traditional A-, B-, and C-kills correspond to time to failure criteria of 5, 20, and 40 minutes, respectively. This, of course, also held true for CTIS performance evaluation.

The experimental setups for the tire deflation and CTIS performance work were identical as far as instrumentation and ballistics were concerned. The differences were in the configuration of the BM-21's tire inflation components. In both cases, the front driver's side tire (designated as Tire 1) was used as the target tire. It was decided that the target tire would remain stationary during the firing process for each experiment since Bodt and Schall (1991) showed that motion of a tire was insignificant and otherwise would require a rather sophisticated setup. Furthermore, the vehicle had to remain stationary during the tire deflation time to allow for collection of pressure-time data. A pressure gauge was installed in line with the air hose connected to Tire 1. This allowed for monitoring of the tire pressure before and during experimentation. Guns ranging in size from 5.56 mm to 12.7 mm were used to fire fragments and fragment simulators ranging in size from 0.13 g to 13.4 g. The setup also contained a stripper plate for the plastic sabots used to hold the fragments, two "sky screens" to measure velocity, and a steel barrier placed behind the front tire to protect engine components of the BM-21 rocket launcher. Prior to conducting each tire deflation experiment, the CTIS was activated to inflate Tire 1 to approximately 3.2 kg/cm^2 (45 psi). The valve to Tire 1 located in the truck cab was then closed to isolate the tire from the rest of the inflation system. This was done to allow the tire to deflate as if it were on a vehicle that did not have a CTIS.

The evaluation of the CTIS performance required a different valve configuration and required that the engine be running throughout each experiment. All tire valves remained in the open position for each CTIS experiment. The CTIS was configured in this manner to represent the CTIS design currently in use. This effectively meant that all of the tires could deflate if a single tire were perforated.

5.2 TIRE DEFLATION ANALYSIS AND RESULTS

The single-tire deflation experiments were conducted to determine the validity of the functions that were being used for tires for describing the probability of component dysfunction given a hit (P_{cdlh}) at the 40-minute time criterion. These functions are provided as inputs to vulnerability codes for use in determining the probability of causing component dysfunction given a hit by a fragment of a certain size and velocity. If the existing functions were found to be invalid, new functions or models were to be developed that could be implemented in vulnerability codes. The P_{cdlh} functions are "step functions" that correlate fragment mass–velocity combinations to a P_{cdlh} value. Figure 1 graphically represents one of the two-step P_{cdlh} functions. "Two-step" means that for each mass, there are two velocity steps that give different P_{cdlh} values. Note that the P_{cdlh} values provided are for single fragment impact on a single tire. There is no ability to account for the effect of multiple fragment impacts. Also note that this function is indicating that a single fragment as small as 0.06 g (1 gr) has the potential to cause tire dysfunction. It is easy to understand why tires have been shown as being quite vulnerable in many vulnerability analyses.

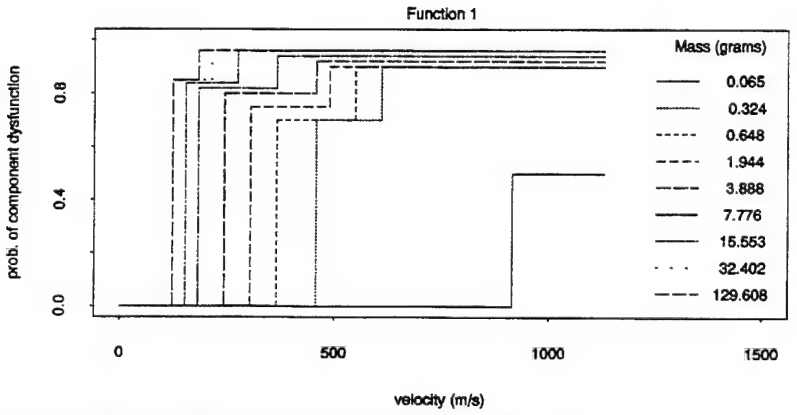


Figure 2. An Example of a Step Function for the Probability of Component Dysfunction Given a Hit.

The first assessment of the experimental data collected reveals that the existing functions dramatically overestimate the vulnerability of tires to single fragment attacks. A few experimental results are provided below to illustrate the point about the inadequacy of the existing $P_{cd/h}$ functions:

1) It took five to six perforations, of a single 14.00-20 tire, by 0.26-g (4 gr) steel fragment simulators to deflate the tire within 40 minutes.

2) 25–26 perforations by 0.26-g (4 gr) steel fragment simulators were required for deflation within 5 minutes.

Thus, it was clear that not only were the P_{cdlh} functions invalid, but a completely new approach would be needed to account for the effects from multiple fragment perforations. The new approach was to develop a model, utilizing regression analysis of the experimental data and engineering calculations, that would calculate tire pressure as a function of time, number of perforations, and fragment size.

Starting with the ideal gas law and experimental results on gas effusion, Equation 5 can be derived (see Grote et al. [1996] for derivation) as

$$P(t) = P(0)\, e^{[\,t\,(\text{constant}\cdot Hn)\,(R\cdot T/V)]} \tag{5}$$

where
- P = tire pressure (kg/cm^2)
- $P(0)$ = initial tire pressure (kg/cm^2)
- t = time (s)
- n = number of perforations
- H = area of a hole (cm^2) (assumes equal size holes)
- T = absolute temperature (K)
- R = universal gas constant (84.73 kg ·cm/ mole ·K)
- V = total volume of system (cm^3).

The model (5) is based on ideal conditions uncomplicated by irregularly shaped holes of varying depths which can change the nature of the air flow. Experiments do not provide a direct measure of the hole area or shape characteristics. To make a link from the shot conditions of the experiments to the functional description of the tire pressure, a slightly altered version of Equation 5 is used, incorporating the fragment presented area into a new constant C:

$$P(t) = P(0)\, e^{-t\,(C)\,(R\cdot T/V)}, \tag{6}$$

where
$$C = \{\, 0.10205\,(\Sigma P_a) + 0.19662\,\Sigma(P_a)/n\}^2$$
and
P_a = presented area of a fragment at impact.

C is not a function of the tire volume (V), the time (t) of the measurement, the initial tire pressure, or temperature. It is the only degree of freedom in Equation 6 for fitting to experimental data, and it is an implicit function of variables related to hole geometry, number of holes, exposed area, and possibly other damage characteristics.

The empirical value for C starts with the basic exponential equation: $P_i(t) = P(0)\, e^{(-t\,\cdot\,Ci)}$. For each of the $i = 58$ data sets, the C_i parameter was determined from a regression in logarithmic form: $\ln(P_i(t)) = \ln(P(0)) - t\cdot C_i$. The R^2 for each was at least 0.95. A common C was determined for the entire data set based on tire damage characteristics. Mass and shape of fragment do not characterize the damage, but instead the fragment itself. However, the average fragment presented area and the cumulative presented area on the tires do characterize damage. Both are statistically significant in a multiple regression analysis and were included as independent variables. The overall fit for C, which is forced through the origin, has an $R^2 = 0.957$. Figure 3 shows the fit of the Tire Deflation Model (Equation 6) for four fragments, 0.26 g (4 gr) each, perforating the tire. The three replications of the experiment illustrate the amount of experimentwise variability in the deflation rate.

The data upon which the tire deflation model was based were for fragments of identical size. Experimental excursions were conducted, showing the adequacy of the deflation model for tires of various sizes, for deflation caused by perforation with multiple fragments of various sizes, and for perforation by much larger fragments. A comprehensive discussion of these excursions is presented in Grote et al. (1996).

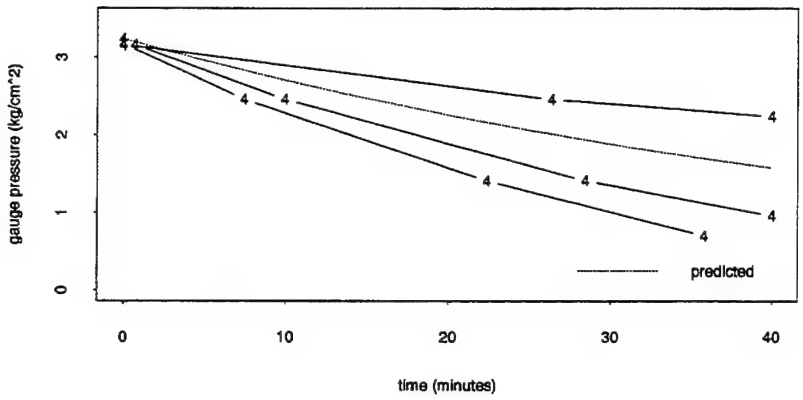


Figure 3. Example of Tire Deflation Fit for Four Fragments, 0.26 g Each.

5.3 CTIS PERFORMANCE ANALYSIS AND RESULTS

The capability of an on-board CTIS has never been explicitly modeled in a vulnerability analysis. Thus, additional experiments were conducted to determine the effect a CTIS could have on the ability of fragments to deflate some or all of the tires on the BM-21. After an appropriate engine speed was determined, 45 experimental firings were conducted. Some interesting results that point out the significance of a CTIS are as follows:

1) The CTIS was able to maintain the maximum tire pressure in all six tires of the BM-21 when a single tire was perforated 35 times by 0.26-g (4 gr) steel RCC fragment simulators.

2) Eleven perforations of a single tire, with 0.97-g (15 gr) steel fragment simulators, caused pressure in all tires to drop to 2.46 kg/cm^2 (35 psi) in 1,227 s (~20 minutes) and 2.11 kg/cm^2 (30 psi) in 2,400 s (40 minutes).

3) A single perforation of a tire by a 13.4-g (207 gr) steel fragment simulator, fired from a 0.50-caliber gun, resulted in a rapid drop in pressure that leveled off at about 2.5 kg/cm^2 (38.6 psi).

Experimentation never proceeded beyond the number of perforations mentioned in 2 and 3 above because once the tires were deflated with those numbers of holes, they could not be reinflated by the CTIS. Since sufficient experimental data could not be obtained to develop an empirical model for CTIS performance, theoretical models were sought.

As a first approximation of the CTIS system, the tires were all assumed to be instantaneously in equilibrium with each other, and one pressure function sufficient for all tires. In effect, there would be one large tire with a volume equal to the sum of the volumes of the individual tires. The reservoir tank, which is generally held at ~7.1 kg/cm^2 (100 psi) prior to a leak, represented another volume. When the CTIS is set in the inflate mode and the engine speed is constant, the compressor provides a constant number of molecules per unit time to the reservoir.

A reservoir tank for both this model and subsequent more complex models appeared to be unnecessary, since the higher starting pressure almost instantaneously equilibrated with the tire pressures. Inclusion of the tank might be needed if the volume in the tank were not such a small fraction of the tire volumes or if the tank pressure were allowed to rise far beyond a fixed cut-off pressure. Without the reservoir, the air from the compressor is assumed to move directly into the tire compartment.

This simplified model yields a closed-form solution:

$$P(t) = m_1/c - (m_1/c - P(0))\, e^{-t(c \cdot R \cdot T / V_t)}, \qquad (7)$$

where
- P = tire pressure (kg/cm^2), gauge pressure
- m_1 = number of moles/time from compressor
- $P(0)$ = initial tire atmospheric (gauge) pressure (kg/cm^2)
- c = rate constant for air leaking from tires to atmosphere (and is estimated by the regression C-value)
- t = time (s)
- T = temperature (K)
- V_t = volume of tire (cm^3)
- R = universal gas constant (84.73 kg ·cm/ mole ·K).

As long as the leak rate of each tire is small or the damage to each tire is essentially the same, this simplified form gives very satisfactory results. However, when some tires have many holes and some have few or no holes, the individual

tires can have quite different pressure histories. Note that if the flow rate from the CTIS (m_1) is set to zero in Equation 7, it can be reduced to the Tire Deflation Model. A complete derivation is omitted for brevity.

An extensive effort was put forth in an attempt to try to preserve the simplicity of this closed-form model. We believed the less information required about a target vehicle, the easier implementation would be in vulnerability codes. Selective application of the simplified CTIS model to groups of tires with similar damage and the Tire Deflation Model to tires with extensive damage could not capture the complexity of a flow in which the CTIS compressor air is dynamically allocated according to relative pressures in the tires.

A complex model having separate air lines from each tire to a common junction which then connected to the compressor was proposed. The model was simplified a bit by assuming the hoses connecting the tires to the common junction were all of the same length and had the same flow characteristics. The value of the constant that characterized these hoses was determined by fitting the model to data for which one, two, and three hoses had been disconnected from the tires.

This model leads to a system of seven coupled, first-order, linear differential equations. A closed-form solution is a complicated sum of seven exponential terms and provides no more insight than a solution obtained by numerical methods. The software developed used a simple Runge-Kutta technique to generate pressure histories for each of the seven functions. Systems with a different number of tires can be used in the program by changing the number of tires in the input

$$P_i'(t) = (RT/V_i)\,[c_h(P_0(t) - P_i(t)) - c_i\,P_i(t)\,], \qquad i = 1, \ldots, n \tag{8}$$

$$P_0'(t) = (RT/V_0)\,[m_1 - c_0 P_0(t) - \Sigma c_i\,(P_0(t) - Pi(t))\,], \quad i = 1, \ldots, n. \tag{9}$$

To use this model, the analyst must be able to provide the number of tires connected to the CTIS, the initial tire pressures, the volume of each tire, deflation constants from the regression formula for each tire, a constant used in characterizing the flow rate from tires to the junction, the compressor capacity in moles of gas per second, the air reservoir volume (not critical — can be approximated), the air temperature, and the atmospheric pressure.

6. SUMMARY AND CONCLUSIONS

The ability of small steel and tungsten fragments to perforate tires was fully characterized, for a wide range of tire thicknesses, via the ballistic limit characterizations and the development of a residual velocity penetration equation. A residual mass algorithm was not developed since there appeared to be no erosion of the fragments upon perforation of the tire targets. Small excursions were conducted to determine how target obliquity and multiple target barriers should be handled when applying the residual velocity algorithm.

The issues of tire deflation and the effect of central tire inflation systems were addressed through the development of three models. First, in the absence of a CTIS, Equation 6 should be applied to each tire individually. This equation was developed via theoretical derivation and regression analysis of experimental data. When the vulnerability of a vehicle that has a CTIS is to be analyzed, the complex CTIS model presented in section 5 should be implemented. The simplified model, Equation 7, provides a good representation when the tire damage is minimal or all tires are damaged to "nearly" the same degree. It is a very difficult task to determine how "nearly" should be defined. Additionally, it is difficult to deflate all tires within standard time criteria of 5, 20, and 40 minutes when minimal damage has occurred.

An issue that was not explicitly stated in the introduction, is the extensibility of the work conducted as part of this effort to vehicles other than the BM-21 and the URAL-375D. Both the tire deflation and the CTIS models are in a general form that allows for application to any other vehicle. To apply these models to other vehicles, certain parameters about the vehicles must be known. For the tire model in Equation 6, the volume of each tire and the initial tire pressure must be known. Of course, a minimum tire pressure must also be provided to determine whether a particular tire would be considered nonfunctional. Additionally, the CTIS model requires the number of tires connected to the CTIS, an airflow rate from the compressor and air tanks to the system of tires, and information concerning the connecting hoses.

Another issue for extensibility, which is not so obvious, concerns the use of the penetration equation that was developed. Tire thickness is a major parameter in the penetration equation, yet most target descriptions have tires with uniform thickness profiles. The differences in tire thickness were considerable over the tread and sidewall areas of the BM-21 tires, not to mention variability that exists from tire to tire. This leads one to believe that to accurately apply the penetration equation to any tire, the thickness profile of the sidewall and tread areas will be required. This further implies that the geometric target descriptions of tires will have to be more detailed than those that currently exist.

7. REFERENCES

Bely, D., B. A. Bodt, and R. N. Schumacher. "Real Bomblet Penetration Evaluation," BRL-TR-3400, U.S. Army Ballistic Research Laboratory, Aberdeen Proving Ground, MD, September 1992.

Bodt, B., and J. Schall. "Survivable Tire System (STS) Test Analysis: Stage 1 Survivability," BRL-TR-3226, U.S. Army Ballistic Research Laboratory, Aberdeen Proving Ground, MD, 1991.

Chambers, J. M., and T. J. Hastie. "Statistical Models in S," Wadsworth & Brooks/Cole Advanced Books & Software, AT&T Bell Laboratories, Pacific Grove, CA, 1992.

Danish, M. B. "Perforation of Truck Tires by Compact Steel Fragments," BRL-MR-1926, U.S. Army Ballistic Research Laboratory, Aberdeen Proving Ground, MD, 1968.

Danish, M. B. "Perforation of Radial Truck Tires by Compact Steel Fragments," BRL-MR-2269, U.S. Army Ballistic Research Laboratory, Aberdeen Proving Ground, MD, 1973.

Draper, N., and H. Smith. Applied Regression Analysis, Second Edition, John Wiley & Sons, Inc, NY, 1991.

Grote, R. L., L. L. C. Moss, and E. O. Davisson. "New Tire Penetration and Pressure Models for Improved Vulnerabilility Analyses of Wheeled Vehicles," ARL report in publication, 1996.

Joint Technical Coordinating Group for Munitions Effectiveness. "Joint Munitions Effectiveness Manual-Surface to Surface, Fragmentation and Terminal Effects Data for Surface to Surface Weapons," JTCG/ME-61S1-3-4, Revision 1, July 1982.

McKaig, A., and J. Thomas. "Maximum Likelihood Program for Sequential Testing Documentation," BRL-TR-2481, U.S. Army Ballistic Research Laboratory, Aberdeen Proving Ground, MD, 1983.

Myers, R. H. Classical and Modern Regression with Applications, Second Edition, PWS-Kent Publishing Company, Boston, MA, 1990.

The Johns Hopkins University. "The Resistance of Various Metallic Materials to Perforation by Steel Fragments; Empirical Relationships for Fragment Residual Velocity and Residual Weight," Project THOR Technical Report, No. 47, for the Ballistic Analysis Laboratory Institute for Cooperative Research, April, 1961.

U.S. Army Materiel Command. "Experimental Statistics - Section 2 - Analysis of Enumerative and Classificatory Data," AMC Pamphlet 706-111, 1969.

U.S. Army Materiel Development and Readiness Command. "Selected Topics in Experimental Statistics With Army Applications," DARCOM Pamphlet 706-103, 1983.

USSR. "Model URAL-375D Truck," Service Manual Seventh Edition, USSR, Moscow, undated.

Warner, D. R. "Ground Transport Vehicles (Current and Projected)—Eurasian Communist Countries," DST-1150S-280-87, Foreign Science and Technology Center, Charlottesville, VA, 1987.

USING REAL-WORLD AND SIMULATION DATA TO ESTIMATE A LOCATION PARAMETER

Dennis E. Smith
Desmatics, Inc.
P.O. Box 618
State College, PA 16804

ABSTRACT

This paper addresses the problem of estimating a location parameter μ using observations from both a real-world system and from a simulation model of that system in the estimation process. The paper examines, for a simple situation, possible estimation methods and investigates how bias in the simulation model affects its usefulness. Even for the "nice" case considered, one cannot be casual about how to use the two types of data.

INTRODUCTION

Simulation has become an indispensable technique for examining complex real-world systems or processes. Since the advent of the digital computer, myriads of simulation models have been constructed. An important question is whether a simulation model is valid, i.e., whether it adequately represents the real-world system it purports to emulate.

When real-world data is available (or can be obtained), simulation validation is often approached via hypothesis testing; the question to be answered is whether or not the simulation is "accurate." However, a hypothesis-testing approach will tend to result in many more inaccurate models being "accepted" than accurate models being rejected, because of the usual imbalance between the probabilities of Type I and Type II errors. This situation can be ameliorated to some extent by consideration of the operating characteristic (OC) curve associated with any test. The OC curve can be used to make more reasonable trade-offs between the two types of errors. An approach that provides more information than one based on a hypothesis test examines confidence intervals on the differences between simulation parameters and those of the corresponding process being modeled. (For a further discussion of simulation validation and references to the literature, see Law and Kelton[1].)

In any event, even if a simulation model is not a faithful representation of the real-world system, that model may still provide useful information. The question then becomes one of how best to use that information in light of the real-world observations. For example, possible answers to this question might be that the simulation data should:

(1) be used as is,
(2) be discarded,
or (3) be modified in some defined manner.

PROBLEM SCOPE

This paper addresses the question of how best to combine two sets of data, one from a simulation model and the other from the corresponding real-world system, when the mean of a univariate response is to be estimated. The particular situation considered is one in which:

(1) n independent real-world observations $y_1, \ldots, y_n$ from $N(\mu,\sigma^2)$ are available

and (2) m independent simulation observations $w_1, \ldots, w_m$ from $N(\mu+\gamma\sigma,\sigma^2)$ are also available.

Of course, the values of μ, σ, and γ are unknown. If $\gamma \neq 0$, the simulation data contains a bias.

Approved for public release; distribution is unlimited.

This situation is essentially a slippage model, which is often used in the investigation of possible spurious data. (See Barnett and Lewis[2], for example.) Although this problem framework could be generalized (e.g., to unequal variances and correlated observations), an objective of this paper is to show that even for this simple and somewhat unrealistic situation, there are still concerns.

This paper examines the usefulness of the simulated data in estimating the location parameter μ as a function of the simulation bias γ, adopting mean square error (MSE) as the criterion for evaluation. If there were no simulation bias, then the maximum likelihood estimator, the pooled mean $(n\bar{y}+m\bar{w})/(n+m)$, would have minimum MSE. Since the assumption of no bias is at best tenuous, this paper examines what happens to this estimator and three additional ones for values of $\gamma \neq 0$.

ESTIMATION OF μ

Suppose it is decided to estimate μ by using an estimator

$$\hat{\mu}_p = p\bar{y} + (1-p)\bar{w}$$

which, as a weighted average of the real-world and simulation observations, is of the same general form as the pooled mean in which each observation is weighted equally, i.e., with $p = p^* = n/(n+m)$.

For general p, where p is a constant,

$$MSE(\hat{\mu}_p) = p^2\sigma^2/n + (1-p)^2[(\sigma^2/m) + \gamma^2\sigma^2] \qquad (1)$$

If only the real-world data is used (p=1), the estimator would be $\hat{\mu}_1 = \bar{y}$. For this estimator,

$$MSE(\hat{\mu}_1) = \sigma^2/n \qquad (2)$$

On the other hand, if both data sources were used, giving each observation equal weight, the estimator would be the pooled mean

$$\hat{\mu}_{p^*} = (n\bar{y}+m\bar{w})/(n+m),$$

for which

$$MSE(\hat{\mu}_{p^*}) = \sigma^2(n+m+m^2\gamma^2)/(n+m)^2 \qquad (3)$$

As one would expect, for values of γ close to zero, the estimator $\hat{\mu}_{p^*}$ based on both sets of observations provides a smaller MSE than that resulting from the use of only the real-world data. It can be seen from (2) and (3) that the use of $\hat{\mu}_{p^*}$ provides smaller MSE so long as $|\gamma| < [(n+m)/nm]^{1/2}$. For larger values of $|\gamma|$, the inflation in MSE rapidly becomes catastrophic; the MSE is unbounded as $|\gamma| \to \infty$.

Such catastrophic results could be avoided by never using the simulation observations, i.e., by always using the estimator $\hat{\mu}_1 = \bar{y}$. However, such a strategy ignores the opportunity to use the simulation data to obtain better estimates when γ is small.

Another approach to this problem is to return to the validation framework and use the pooled estimator $\hat{\mu}_{p^*}$ if the simulation model is judged valid, or use the estimator $\hat{\mu}_1$ otherwise. An appropriate hypothesis test for testing model validity would involve the hypotheses

$$H_0: \gamma = 0$$
$$H_1: \gamma \neq 0$$

based on the t-statistic

$$t = [nm/(n+m)]^{1/2}(\bar{y}-\bar{w})/s$$

where s denotes the pooled estimated standard deviation. The critical region, assuming a significance level of α, would be

$$|t| > t(\alpha/2,n+m-2)$$

where $t(\alpha/2,n+m-2)$ denotes the upper $\alpha/2$ point of a t-distribution with n+m-2 degrees of freedom.

If H_0 were rejected, the estimator $\hat{\mu}_1 = \bar{y}$ would be used, while if it were not rejected, the estimator $\hat{\mu}_{p=} = (n\bar{y}+m\bar{w})/(n+m)$ would be used. An estimate of γ is given by

$$\hat{\gamma} = (\bar{y}-\bar{w})/s$$

Therefore, the validation approach results in the use of the estimator

$$\hat{\mu} = \begin{cases} \hat{\mu}_{p=} & \text{if } |\hat{\gamma}| < [(n+m)/nm]^{1/2}t(\alpha/2,n+m-2) \\ \hat{\mu}_1 & \text{if } |\hat{\gamma}| \geq [(n+m)/nm]^{1/2}t(\alpha/2,n+m-2) \end{cases}$$

where the value of α must be specified.

This estimator is based on an all-or-nothing rule whereby if $\hat{\gamma}$ is too large, only the real-world data is used, and if $\hat{\gamma}$ is not too large, all observations are used. Thus, a simulation observation is given either a weight of zero or equal weight with each real-world observation, depending upon the size of $\hat{\gamma}$.

A more flexible procedure would incorporate $\hat{\gamma}$ directly into an adaptive estimator. To arrive at a reasonable adaptive estimator, note from (1) that by setting

$$\partial MSE(\hat{\mu}_p)/\partial p = 0,$$

one finds that

$$p = (n+nm\gamma^2)/(n+m+nm\gamma^2) \qquad (4)$$

provides the minimum MSE. If $\gamma = 0$, $p = n/(n+m)$, so that $\hat{\mu}_p$ reduces to the weighted average with each real-world and simulation observation given the same weight.

Because γ is an unknown parameter, the value of p providing the minimum MSE is also unknown. Thus, one might consider substituting $\hat{\gamma}$ for γ in (4) and using the resulting value as an estimate of p. This procedure results in an adaptive estimator $\hat{\mu}_{\hat{p}}$. Because $\hat{p}$ is a random variable rather than a constant, $MSE(\hat{\mu}_{\hat{p}})$ cannot be obtained by substitution into (1).

EVALUATION OF THE ESTIMATORS

In this paper, four possible estimators have been considered for the task of estimating the location parameter μ with real-world and simulation data. These estimators are:

(1) $\hat{\mu}_{p=}$, which always uses the real-world and simulation observations weighted equally,

(2) $\hat{\mu}_1$, which always uses only the real-world observations,

(3) $\hat{\mu}$, which is based on a hypothesis test of the validity of the simulation,

and (4) $\hat{\mu}_{\hat{p}}$ which is an adaptive estimator based on an estimate of γ.

The performance of these estimators does not depend upon the actual values of μ and σ since the bias of any simulation observation is measured in units of σ and location has no effect on the results.

To compare the performance of these four estimators for a set of n real-world and m simulation observations, their MSEs must be evaluated for a range of γ values. This poses no difficulty in the case of the first two estimators; the required MSEs are given by (1) and (2). Unfortunately, things are not so easy for the other two estimators. For the MSE of $\hat{\mu}$ one must compute, in $(s, \bar{y}, \bar{w})$ space, the expected value of $[(n\bar{y}+m\bar{w})/(n+m)-\mu]^2$ over the region

$$|(\bar{y}-\bar{w})/s| < [(n+m)/nm]^{1/2}t(\alpha/2,n+m-2)$$

and the expected value of $(\bar{y}-\mu)^2$ over the region

$$|(\bar{y}-\bar{w})/s| \geq [(n+m)/nm]^{1/2}t(\alpha/2,n+m-2)\ .$$

For the MSE of the adaptive estimator $\hat{\mu}_{\hat{p}}$, one must evaluate the expected value of

$$[\{n+nm[(\bar{y}-\bar{w})/s]^2\}\bar{y}+m\bar{w}]/[\{n+m+nm[(\bar{y}-\bar{w})/s]^2\}-\mu]^2.$$

Because of their complexity, an analytic evaluation of these expected values is an impossible task; Monte Carlo was used to evaluate the MSEs.

Performance of the four estimators was evaluated for two cases: n=3, m=10 and n=3, m=50. For the validity test estimator $\hat{\mu}$, four values of α (.01, .05, .10, and .20) were considered. Table 1 and 2 list the MSEs of the four estimators for these cases relative to the MSE of $\hat{\mu}_1 = \bar{y}$, which is $\sigma^2/3$ in both cases considered. As can be seen, none of the four estimators dominates or is dominated by any other estimator.

SUMMARY

It is clear that the pooled mean estimator $\hat{\mu}_{p^*}$, with its unbounded MSE, is not worth considering. By using either the validity test estimator $\hat{\mu}$ or the adaptive estimator $\hat{\mu}_{\hat{p}}$, one will come out ahead, or at least not too far behind, if $|\gamma|$ is either small or large. It is for moderate values of $|\gamma|$, approximately $1.0 < |\gamma|/\sigma < 3.0$, that the worst things occur. The adaptive estimator appears to be the rational choice since it provides reasonable gains (decreases in MSE) for small $|\gamma|$ and, in the worst case, does not result in a substantial penalty.

These results indicate that a standard hypothesis test of model validity may be hazardous if the data is to be used for estimation of μ. This can be seen from the tables.

REFERENCES

1. Law, A. and D. Kelton. Simulation Modeling and Analysis (2nd Edition), New York: McGraw-Hill, Inc., 1991.

2. Barnett, V. and T. Lewis. Outliers in Statistical Data (3rd Edition), New York: John Wiley and Sons, Inc., 1994.

Table 1: MSEs of Estimators Relative to $MSE(\hat{\mu}_1) = MSE(\bar{y})$ for n=3, m=10.

$\lvert\gamma\rvert$	Pooled Mean Estimator $\hat{\mu}_{p\bullet}$	Validity Test Estimator $\hat{\mu}$				Adaptive Estimator $\hat{\mu}_{\hat{p}}$
		α=.01	α=.05	α=.10	α=.20	
0.0	0.23	0.29	0.39	0.50	0.66	0.62
0.5	0.68	0.77	0.91	0.97	0.99	0.75
1.0	2.01	2.02	1.75	1.58	1.32	1.07
1.5	4.22	3.34	2.27	1.82	1.38	1.20
2.0	7.33	3.96	2.19	1.60	1.25	1.17
3.0	16.21	2.80	1.32	1.17	1.13	1.11
4.0	28.63	1.28	1.15	1.13	1.10	1.05
5.0	44.61	1.09	1.07	1.05	1.04	1.04

Table 2: MSEs of Estimators Relative to $MSE(\hat{\mu}_1) = MSE(\bar{y})$ for n=3, m=50.

$\lvert\gamma\rvert$	Pooled Mean Estimator $\hat{\mu}_{p\bullet}$	Validity Test Estimator $\hat{\mu}$				Adaptive Estimator $\hat{\mu}_{\hat{p}}$
		α=.01	α=.05	α=.10	α=.20	
0.0	0.06	0.12	0.27	0.43	0.64	0.48
0.5	0.72	0.88	0.98	1.04	1.06	0.73
1.0	2.73	2.63	2.16	1.87	1.54	1.11
1.5	6.06	3.91	2.53	1.95	1.44	1.20
2.0	10.74	3.08	1.70	1.31	1.18	1.26
3.0	24.09	1.00	1.00	1.01	1.17	1.07
4.0	42.78	0.99	0.99	0.99	1.00	1.05
5.0	66.81	0.97	0.98	0.98	1.01	1.03

INTENTIONALLY LEFT BLANK.

THE EFFECTS OF A COMPUTER-AIDED TELEOPERATION TECHNOLOGY ON OPERATOR WORKLOAD AND PERFORMANCE OF CONCURRENT TASKS

Monica M. Glumm and Jock O. Grynovicki
U. S. Army Research Laboratory
Human Research and Engineering Directorate
Aberdeen Proving Ground, MD 21005-5425

ABSTRACT

The feedback limited control system (FELICS) is a computer-aided teleoperation (CAT) technology that enables the remote operator to designate an extended path that the vehicle will automatically follow. This paper describes the methodology and results of a study designed to quantify the effects of this technology on remote driving performance and operator workload in both single and dual task conditions. In the dual task condition, the operator's ability to detect and identify targets while driving was also measured. These data were compared with those obtained when the same vehicle was operated in the standard mode of remote driving.

Generally, the findings indicate that operators in the CAT mode did not attain the speeds and committed more driving errors than operators in the standard mode ($p < .001$). In the CAT mode, operators rated the effort they expended to achieve their level of performance higher than did those in the standard mode ($p < .05$). In the dual task condition, driving errors increased in the CAT mode ($p < .05$) and fewer targets were correctly identified than in the standard mode of remote operation ($p < .01$).

INTRODUCTION

In both the computer-aided and standard modes of remote driving, the operator's task is to designate the vehicle's path. In the standard mode, the operator maneuvers the vehicle through the scene displayed on a video monitor, providing continuous control input to which the vehicle responds in near real time. In the computer-aided mode, the operator plots an extended path within the driving scene which the vehicle will automatically follow. In this mode, while the vehicle is maneuvering along the designated path, the role of the driver is more that of a supervisor. During this interval in time, the remote driver monitors the progress of the vehicle and watches for any hazards that may not have been detectable from previous positions. This technique of remote driving theoretically offers a reduction in operator workload and potentially enables simultaneous control of another vehicle or the performance of another task. Additionally, in this mode, driver effectiveness may possibly be sustained at video update rates far below those required to control a vehicle in the standard mode of remote operation. This

---

[1] Measured using the Snellen visual acuity chart.

latter capability could result in significant reductions in communications bandwidth and enhance vehicle survivability on the battlefield.

Although the concept of computer-aided teleoperation is not a new one, there has been little research that supports the anticipated benefits of the concept or that might assist in the development of a technology that will. This paper presents the results of a study that may provide insight into the design and training issues that challenge the developers of this concept -- issues that must be resolved before some of the possible benefits of this new technology can be realized.

The computer-aided teleoperation (CAT) technology assessed during this investigation was the feedback limited control system (FELICS), developed by AmDyne Corporation of Millersville, MD. An initial demonstrator was built under a Phase 1 Small Business Innovative Research (SBIR) contract with the Human Research & Engineering Directorate (HRED) of the U.S. Army Research Laboratory (ARL). Further development of this system was funded by the program manager-unmanned ground vehicles (PM-UGV) at Redstone Arsenal, Alabama. The present study, which was conducted at the request of the PM-UGV, attempted to quantify the effects of the FELICS on remote driving performance and operator workload during both single and dual task conditions. In this study, a reduction in the subjects' experiences of workload was expected to be reflected in an increase in the operators' ability to drive and detect and identify targets concurrently.

METHOD

SUBJECTS

The 32 military volunteers who participated in this study were licensed drivers between the age of 19 and 34 years. All were screened to ensure color vision and visual acuity of 20/20 vision in one eye and at least 20/100 in the other eye (corrected or uncorrected).

APPARATUS

Research Platform and Control Stations. The research platform was a four-wheel, electrical golf cart, converted by the designer of FELICS to enable operation in either the computer-aided or the standard mode of remote driving. The control station used for driving the vehicle in the standard mode consisted of a steering wheel, brake, and accelerator pedal. In the CAT mode, a displacement joystick with knob controlled both the direction and steering of the cursor that spawned waypoints, indicating the vehicle's future path. A forward movement of the joystick advanced the cursor in the forward direction. Turning the knob at the top of the joystick to either the left or right

steered the cursor. A rearward movement of the joystick enabled the operator to withdraw some or all of the waypoints plotted. If all waypoints were withdrawn, the vehicle was stopped. This control design had been selected during an earlier pilot study from two other candidate controllers, which included the standard controls and a two-axis force stick that was supplied by the contractor.

In the CAT mode, the vehicle's ability to attain maximum speed was determined by the straightness and the length of the path (number of waypoints) that the operator had plotted. The maximum number of waypoints that could be laid on the course at any one time was restricted to 15 by the contractor to minimize down-range and cross-range errors associated with vehicle execution of the designated path. For each of the 15 waypoints attained, another could be plotted. Each pair of waypoints was 1 m apart. Thus, the maximum length path that could be plotted at any one time was approximately 15 m (49 ft).

Video Camera and Monitors. For each mode of operation, the video image was supplied by a 1/2-inch charged couple device (CCD) color camera mounted on board the remote platform. A 6-mm focal length lens provided the remote operator an approximate 55° horizontal and 43° vertical field of view (FOV). The driving scene was displayed on a 13-inch color monitor. The terrain scenes and targets were displayed on three 20-inch color TV monitors above the driving display. The resolution of the camera, lens and display assembly was 20/200 [1]. A different camera location was selected for each mode of remote driving to accommodate the distinct differences between driving technologies and to avoid biasing performance in one or the other of these modes. The location of the camera used for operations in the CAT mode was the decision of the contractor who designed the system. This camera was mounted on the left side of a pan-tilt mechanism that was centered laterally on the vehicle. The camera used for operations in the standard mode was fixed approximately 0.8 m (2.6 ft) below the pan-tilt device.

Test Course. The study was conducted on an indoor test course where driving speed and error are measured automatically. The course consists of five segments that include straightaways, turns (right and left hand), serpentine, figure 8, and an obstacle avoidance segment. For the first four segments of the course, the measure of driving error is the distance traveled off the roadway by one or more of the vehicle's wheels. For the last segment (obstacle avoidance), error is based on the number of obstacles hit. For this study, three-dimensional cloth objects were hung along the roadway of the course to represent trees and shrubs that briefly obscured the remote operator's view of the road ahead.

PROCEDURES

During the study, each of the 32 subjects was randomly assigned to one of two groups. One group (Group A) was trained and tested in the CAT mode using

FELICS, and the other (Group B) was trained and tested in the standard mode of remote driving. Before training in remote driving, all subjects received training in target recognition and identification, as well as instruction in assessing their workload experience using the National Aeronautics and Space Administration-Task Load Index (NASA-TLX). During training in remote driving, each subject made consecutive runs through the test course in the assigned mode of operation until an asymptote had been attained in both driving speed and accuracy. The subject then completed two additional trials in which he or she received practice in performing the target identification and driving tasks concurrently. A total of six test trials was then performed in the assigned mode of remote driving. During three of these trials, the subject's only task was to drive the vehicle; during another three trials, the subject was required to perform the driving and target-identification tasks concurrently. The order of presentation of single and dual task conditions was counterbalanced.

During test, each of the 16 subjects within Group A was shown different target scenarios; however, subjects within Group B were shown the same scenarios as their counterparts in Group A. During any given target scenario, a total of 18 targets was presented. Each of the six target types was presented once on each of the three target monitors in a randomized order. The locations at which these targets appeared in the scene were also randomized. The sizes of the targets varied, based on their location within the scene. Each target was presented one at a time for 3 seconds. Target presentation was effected by microswitches located every 4.9 m (16 ft) along the roadway of the test course. Except for the obstacle avoidance segment, an equal number of targets were presented in each course segment during each target scenario. In any given course segment, those switches that effected target presentation were varied among scenarios. A time-to-target presentation of 1 to 3 seconds was randomly generated when a designated switch was tripped by the wheels of the remote platform. One of the six target types then appeared at one of the six target locations on one of the three monitors in accordance with a pre-determined scenario on file in the computer. When the target was displayed, a counter was activated. When the subject detected and announced "target," the investigator immediately depressed a pushbutton that stopped the counter. Time to detect and vehicle speed at detection were stored. The subject was then required to identify the target by owner and type. The investigator compared the subject's response with the programmed scenario, noting whether the target was correctly identified. After 3 seconds, the target automatically disappeared from the screen.

RESULTS

ERROR

Mean driving error (distance traveled off the road) on the first four segments of the course was subjected to an analysis of covariance (ANCOVA) with driving mode (STD versus CAT) as a between-subjects effect, and task conditions (single versus dual) and course segments as within-subject effects.

A significant main effect was found for driving mode, $F(1,29) = 21.6$, $P < .001$, with mean errors of .21 m and .77 m for the CAT and STD modes, respectively. This main effect was attributed to difficulties that drivers in the CAT mode experienced in judging waypoint positioning at distances from the vehicle. The ANCOVA also revealed a significant main effect for condition, $F(1, 29) = 4.34$, $P < .05$ with mean errors of .77 m and 1.01 m for the single and the dual task conditions, respectively. This finding simply indicates that drivers commit more driving errors when required to perform a second task. The significant effect found for segment, $F(3, 90) = 7.86$, $P < .001$, is primarily attributed to the greater driving accuracy achieved on the less difficult straightaway segments of the course. More importantly, a significant interaction for segment and driving mode, $F(3, 90) = 7.66$, $P < .001$, is shown in Figure 1. This interaction is attributed to the lack of a difference in error between the STD and the CAT mode on straightaways as compared to the large differences in errors that occurred between these modes on the other segments of the course. All other effects failed to reach significance at the .05 level of confidence.

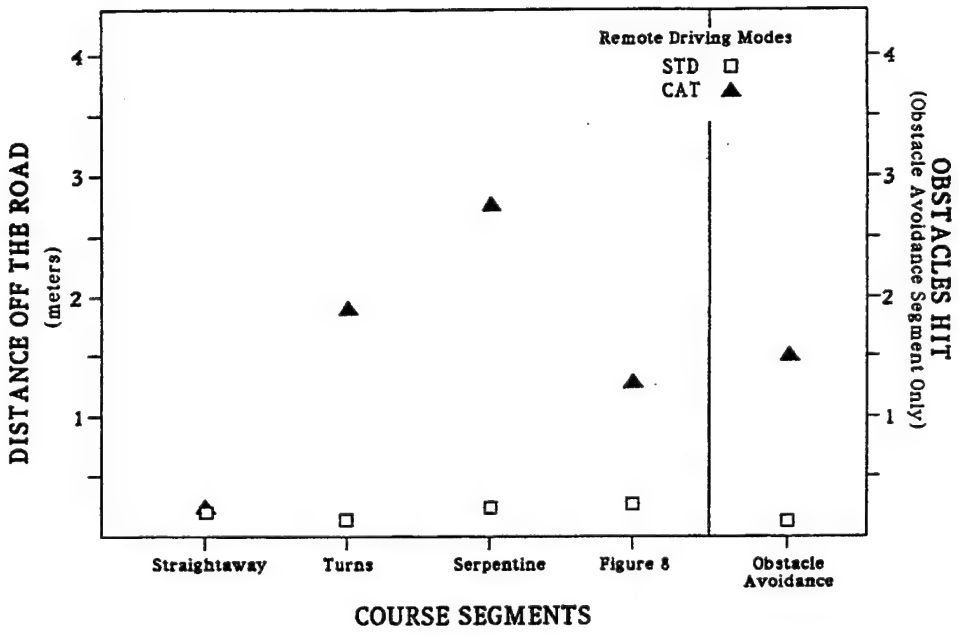


Figure 1. Mean driving error by remote driving mode and course segment averaged over task conditions.

In the last segment of the course (obstacle avoidance) by contrast to the first four segments, the measure of driving error was the number of obstacles hit. Therefore, a separate ANCOVA was performed on this segment with driving mode and conditions as within effects. The analysis revealed a significant main effect for driving mode, $F(1, 29) = 218.21$, $P < .001$, with a mean number of hits of .81 and 1.50 for the STD and CAT modes, respectively. This effect is attributed to the design of the cursor and its offset from the centerline of the vehicle which caused difficulties in judging vehicle position with respect to obstacles. All other effects failed to reach significance at the .05 level of confidence.

SPEED

Mean driving speed on the first four segments of the course was also subjected to an ANCOVA with driving mode (STD versus CAT) as a between-subjects effect, and task conditions (single versus dual) and course segments as within-subject effects. A significant main effect was found for driving mode, $F(1,29) = 196.3$, $P < .001$, with mean speeds of 7.64 kph and 4.69 kph for the STD and CAT modes, respectively. This main effect was attributed to a design feature of CAT that automatically reduces the speed of the vehicle in anticipation of turns to maintain vehicle stability. The ANCOVA also revealed a significant main effect for conditions, $F(1, 29) = 25.3$, $P < .001$, with mean speeds of 6.36 kph and 5.96 kph for the single and the dual task conditions, respectively. This finding simply indicates that drivers drive more slowly when required to perform a second task. The significant effect found for segment, $F(3, 90) = 296.0$, $P < .001$, is primarily attributed to the greater speeds achieved on the straightaway segments of the course by comparison to any other course segment. Also, speeds in the turns were higher than those in the less predictable serpentine. More importantly, a significant interaction for segment and driving mode, $F(3, 90) = 27.81$, $P < .001$, is shown in Figure 2. This interaction is attributed to the somewhat smaller difference in speed between the STD and the CAT mode on straightaways as compared to the other segments of the course. All other effects failed to reach significance at the .05 level of confidence.

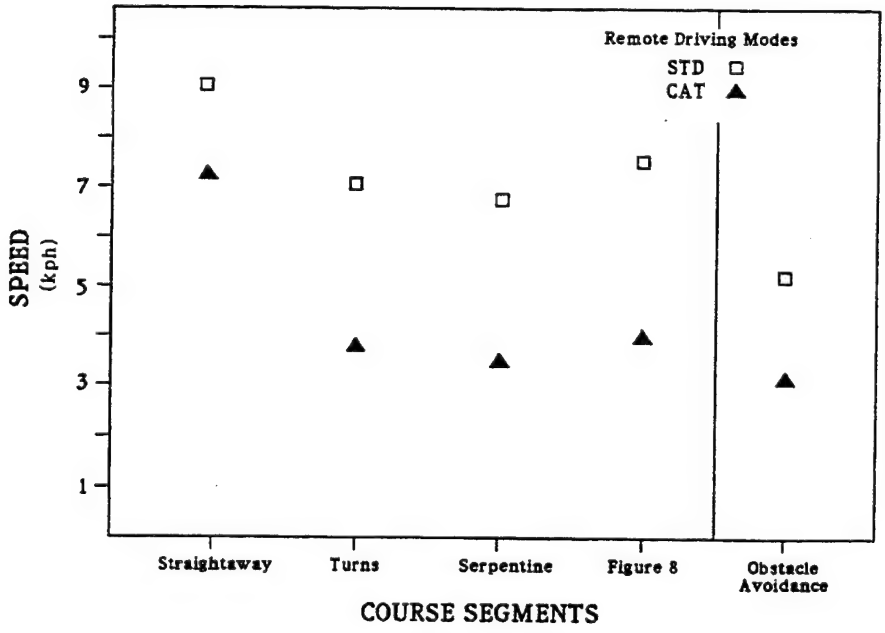


Figure 2. Mean driving speed by remote driving mode and course segment averaged over task conditions.

To be consistent with the analyses of error, a separate ANCOVA was performed on course Segment 5 (obstacle avoidance) with driving mode and conditions as within effects. The analysis revealed a significant main effect

for driving mode, $F(1, 29) = 72.35$, $P < .001$, with mean speeds of 5.21 kph and 3.10 kph for the STD and CAT modes, respectively. As in the analysis of speed on the first four segments of the course, this effect is attributed to a design feature of CAT that reduces the speed of the vehicle in anticipation of turns. All other effects failed to reach significance at the .05 level of confidence.

TARGET IDENTIFICATION PERFORMANCE

The chi-square statistic was used in the analysis of the mean number of targets correctly identified with driving mode (STD versus CAT) a between-subjects effect and course segment a within-subject effect. A significant main effect was found for driving mode, $\chi^2 = 8.28$, $P < .01$, with a mean number of correct identifications of 9.80 and 8.42 for the STD and CAT modes, respectively. This main effect may be attributed to a reduction in the amount of time that subjects in the CAT mode spent in target inspection to confirm their identity. All other effects failed to reach significance at the .05 level of confidence.

The mean time to detect those targets correctly identified and the mean driving speed at the time these targets were detected were each subjected to an analysis of variance (ANOVA) with driving mode (STD versus CAT) a between-subjects effect and course segment a within-subject effect. In the anlaysis of time to detect, all effects failed to reach significance at the .05 level of confidence. However, the results of the ANOVA for driving speed at the time these targets were detected, revealed a significant main effect for driving mode, $F(1,30) = 106.2$, $P < .001$, with a mean speed of 6.21 kph and 3.97 kph for the STD and CAT modes, respectively. As might be expected, a significant main effect was also found for segment, $F(4) = 135.2$, $p < .001$, and there was a significant interaction for segment and driving mode, $F(4, 120) = 8.48$, $P < .001$.

WORKLOAD

The results of the ANCOVAS for speed and error show that there was no relationship between workload and driving speed in either mode of operation, but the subjects' ratings of their performance appeared to be influenced by the distance they traveled off the road, $t = 2.148$, $P < .05$. An association was also found between the subjects' level of frustration and the number of obstacles hit in the obstacle avoidance segment of the course, $t = 2.460$, $P < .05$. A multiple analysis of variance (MANOVA) was performed to determine if there were differences in the subjects' ratings of workload demands between driving modes and task conditions. The results of this MANOVA based on the Wilks statistic indicate that, in the CAT mode, the subjects rated the effort they expended to achieve their level of performance higher than did those subjects who operated the vehicle in the standard mode, $F_{approx} = 4.42$, $P < .05$. The subjects' assessment of their workload in the two driving modes followed similar trends in both task conditions. In the dual task condition, the operators' ratings of mental ($F_{approx} = 9.52$, $P < .05$) and temporal ($F_{approx} = 4.80$, $P < .05$) demands increased significantly in both driving modes.

DISCUSSION

In this study, when subjects were required to perform a second task while driving, driving speed decreased in both modes of operation. In the CAT mode, driving error on some segments of the course increased significantly, but contrary to hypothesis, no relationship was found between these reductions in driving performance and the subjects' ratings of workload. In the dual task condition, driving error in the standard mode was unaffected, and more targets were correctly identified in this mode than in the CAT mode of operation. These findings may reflect, in part, differences in time-sharing efficiency between the two subject groups which to a great extent, as Wickens[1] notes, are related to either differences in the automaticity of single-task skills or to time-sharing skills acquired through practice.

For many, the task of driving a standard automobile over known terrain is somewhat automatic when one arrives at his or her destination with no memory of the trip. Although not quite as automatic, remote driving in the standard mode shared more similarities to the on-board driving experience than did CAT. Commonalities in control design and operation and in the information provided in the driving scene, along with years of familiarization with on-board driving, may have facilitiated time sharing of tasks in the standard mode. Dissimilarities between the on-board driving experience and CAT may have caused some conflict. In the everyday operation of a motor vehicle, a driver engages in visual scanning and cognitive processing activities not unlike those involved in the detection and identification of targets. Wickens suggests that most learned time-sharing skills are probably specific to a given task combination, but one might also expect a transfer of these skills between some task combinations that are similar.

In this study, it was observed that subjects operating in the CAT mode adopted one of three different driving strategies. Some chose to maintain as many waypoints on the course as possible, regardless of segment difficulty, to maximize vehicle speed. Others varied the length of the future path based on segment difficulty and their ability to discern the edges of the road at distances from the vehicle. In a third driving strategy, the subjects chose not to plot an extended path in any segment of the course but rather to maintain a relatively consistent number of waypoints in front of the vehicle. Generally, in this latter strategy, the length of the path or the number of waypoints maintained represented less than half the maximum allowed by the system. It was observed that many of the operators who employed this strategy maintained a fairly consistent speed through most segments of the course, having less need to correct for gross deviations off the road. In some instances, these subjects committed fewer errors and achieved overall course speeds that were similar to operators who adopted more aggressive strategies. The shorter the future path, however, the more closely this driving strategy resembled operations in the standard mode and the more familiar experience of on-board driving.

Generally, those subjects who operated the vehicle in the standard mode of remote driving attained greater speeds with fewer errors than did those subjects who operated the same vehicle in the CAT mode. The significantly lower speeds attained in the CAT mode are believed to be largely attributable to a design feature that automatically reduces the speed of the vehicle in anticipation of turns to maintain vehicle stability. Difficulties in judging waypoint positioning with respect to road edges and obstacles at distances are expected to be a major cause of errors. In many instances, subjects operating in the CAT mode were observed to redraw a path two to three times, to correct for actual or perceived errors in designation as viewed from new and closer camera perspectives. This, too, may have contributed to reductions in vehicle speed and to the higher levels of effort experienced by subjects in this mode. The increase in errors in the dual task condition may not only reflect a decrease in the accuracy of designating the initial path but possibly a reduction in the speed and frequency at which deviations beyond road boundaries were detected and thus, successfully corrected.

Problems in discerning the spacing between obstacles and road edges at distances were compounded by uncertainties about the vehicle's position. In the standard mode, operators were provided a view of the vehicle and ground proximate to the remote platform. In this mode, operators appeared to gauge with accuracy the position of vehicle's wheels with respect to road edges and confidently cut corners in turns in pursuit of the most efficient path through the course. In the CAT mode, however, a view of the vehicle could not be captured within the operator's visual field unless all waypoints were withdrawn and the vehicle stopped. In this mode, operators were forced to estimate vehicle location based on the position of the cursor and the waypoints it spawned. The design of the cursor, however, and its offset to the left of the centerline of the vehicle made such estimates difficult. The cursor did not resemble the remote platform in either shape or size. Operators knew approximately where to position the cursor with respect to the centerline of the road so as to center the vehicle between the road's borders, but they remained uncertain about how far to the left or right of this point they could deviate without overshooting these borders. In the CAT mode, there was a greater tendency for operators to designate a path that closely conformed to the curvature of the road. Waypoints that appeared to deviate from this more reliable track were often withdrawn and the path redesignated. The design and offset of the cursor created particular difficulties in the obstacle avoidance segment of the course where differences in the clearances to the left and right of the vehicle were a major source of confusion and error. In this segment, operators underestimated clearances between the vehicle and traffic cones to the left of the remote platform, causing them to make wider turns around these obstacles. Limitations in the turning radius of the vehicle combined with operator overestimation of the clearance between the vehicle and traffic cones to the right of the remote platform resulted in obstacle hits.

CONCLUSIONS AND RECOMMENDATIONS

In both the standard and computer-aided mode of remote driving, the operator relies on visual information within the scene to select a safe and efficient path. As one might expect, deficits in information that may at times

affect the teleoperator's ability to judge the suitability of more immediate paths, may have an even greater impact on the operator's ability to assess the suitability of distant terrain and designate with accuracy the route selected. System resolution is not thought to have had a significant influence on driving error in the current assessment, but it is expected to be a factor during cross-country travel or in instances when road edges and obstacles are not as well defined.

In this study, differences in driving performance between the two driving modes are believed to have been influenced by design characteristics of the specific CAT system assessed as well as by problems inherent to similar systems and concepts that rely on the remote operator's ability to see and thus accurately designate a suitable path at distances from the vehicle.

In the present assessment, one of the major causes of differences in speed between driving modes is a design feature that automatically reduces the speed of the vehicle in anticipation of future deviations from a straight line path. Although the designer of FELICS believes computer control of vehicle speed necessary to maintain vehicle stability, it is recommended that the system be modified to provide the remote operator the option of assuming responsibility for such decisions, as tactics direct and terrain permits.

In this study, the design of the cursor and its offset from the centerline of the vehicle was a major source of confusion and error. It is recommended that the cursor be redesigned to accurately depict the size and perspective of the vehicle as the distance and the angle from which this cursor is viewed change. The camera should either be centered laterally on the vehicle or corrections made in the programming so that the centerline of the cursor accurately denotes the centerline of the vehicle.

An increase in camera height may improve the operator's perspective of the future path, and a zoom capability may provide some help in detecting hazards and discerning road edges; image stabilization then becomes a necessity. Sensors on board the remote platform may prove useful in providing information about terrain roughness or the proximity of obstacles and other hazards that may not have been detectable from previous positions. Such sensors may be necessary when update rate and resolution are reduced to achieve a reduction in communications bandwidth. Nonetheless, operator perspective and the difficulties it may cause in judging vehicle position with respect to road edges and obstacles at distances will remain a factor that may limit the length of the future path and the accuracy with which it is designated.

REFERENCES

1. Wickens, C. Engineering Psychology and Human Performance. New York: Harper-Collins, 1991.

A MARKOVIAN MODEL FOR GROWTH AND BRANCHING OF STABLE CRACKS IN METALS

Luis A. de Bejar

U.S. Army Engineer Waterways Experiment Station
3909 Halls Ferry Road
Vicksburg, MS 39180-6199

ABSTRACT

Stemming from an experimentally verified power law for the rate of growth of cracks in metals, a probabilistic description of crack propagation in structural members subjected to a cyclic state of tensile stress is developed using the one-dimensional theory of Brownian motion. The second-moment statistical characterization of the size of the crack at a specific future time is presented in closed form under certain assumptions concerning the distribution of the random variables entering the formulation. After reaching a critical size, the crack multiple splitting (proliferation) is described in the light of the theory of Markovian branching processes.

CRACK GROWTH POWER LAW. DISCRETE REPRESENTATION

Based on experimental evidence, Paris and Erdogan[1] proposed that the crack-growth rate in metal components be expressed in terms of the range of stress intensity factor by means of the following power law:

$$\frac{da}{dN} = C\,(\Delta K)^m, \tag{1a}$$

where:

a = instantaneous crack size (The crack size is defined as the crack semi-length for a centrally located through-thickness crack and as the crack length for an edge crack).
N = number of cycles
ΔK = range of stress intensity factor
C, m = experimental coefficients for the specific material being considered.

This power law represents well the scatter in the fatigue crack-growth behavior of metals for the domain of the range of stress-intensity factor between the threshold value (ΔK_{th}) and the region of unstable crack growth[2-3]. The exponent m is frequently taken as deterministic for practical applications and the coefficient C has been found to be log-normally distributed[2]. Expressed analytically, the probability density function of C: $\log N(\mu_c, v_c)$ is given by :

Approved for public release; distribution is unlimited.

$$p_C(C) = \frac{1}{\zeta_C\, C\sqrt{2\pi}} \exp -\frac{1}{2}\left(\frac{\ln C - \lambda_C}{\zeta_C}\right)^2, \qquad 0 < C \leq \infty$$

$$\textit{in which} \quad \lambda_C = \mu_{\ln C} = \ln \mu_C - \frac{1}{2}\zeta_C^2,$$

$$\zeta_C^2 = \sigma_{\ln C}^2 = \ln (1 + v_C^2),$$

and where μ is the mean and v is the coefficient of variation of the variate in the subscript, C in this case.

The range of stress-intensity factor is related to the range of applied stress by the relation[4]:

$$\Delta K = S \sqrt{\pi a}\; F\left(\frac{\pi a}{w}\right), \tag{1b}$$

where S is the range of applied stress, with problem-specific distribution, and w is the specimen width, taken as deterministic. The empirical function F (πa/w) depends on the location of the crack within the specimen and may be approximated by the expressions:

$$F\left(\frac{\pi a}{w}\right) = \sqrt{\sec\left(\frac{\pi a}{w}\right)}, \tag{2a}$$

for a centrally located through-thickness crack[3], and:

$$F\left(\frac{\pi a}{w}\right) = \sqrt{\frac{2w}{\pi a}\tan\left(\frac{\pi a}{2w}\right)}\;\frac{0.752 + 1.286\,\frac{\pi a}{2w} + 0.37\left(1 - \sin\left(\frac{\pi a}{2w}\right)\right)^3}{\cos\left(\frac{\pi a}{2w}\right)}, \tag{2b}$$

for an edge crack[4].

Letting $\xi = \pi$a/w, substituting eq. (1b) into eq. (1a) and grouping ξ-terms on the right-hand side of the equation leads to the expression:

$$\pi C S^m w^{\frac{m-2}{2}} \Delta N = \int_{\xi_0}^{\xi} \left(\frac{1}{z\,[F(z)]^2}\right)^{\frac{m}{2}} dz$$

$$= \int_{\xi_0}^{\xi_0 + \Delta\xi} \left(\frac{1}{z\,[F(z)]^2}\right)^{\frac{m}{2}} dz \qquad (2c)$$

$$\doteq \left(\frac{1}{\xi_0\,[F(\xi_0)]^2}\right)^{\frac{m}{2}} \Delta\xi .$$

This last equation represents the discretized, or finite difference, version of eq. (1a).

In order to simplify the problem as much as possible, the cyclic stress on the component will be assumed to be simple harmonic with known frequency f and random amplitude range S (an ideal narrow-band process). If we examine the effect of this cyclic stress on the crack during the small time interval Δt, then, after substituting $f\,\Delta t$ for ΔN, eq. (2b) gives the following expression for the departure from the initial crack size, a_0 (random):

$$L = a - a_0 = C\left(\sqrt{\pi a_0}\, F\left(\frac{\pi a_0}{w}\right) S\right)^m f\,\Delta t. \qquad (3)$$

PROBABILISTIC DESCRIPTION OF CRACK PROPAGATION

Linearizing eq.(1a) about the mean vector of basic variables, the moments of L may be estimated from[5]:

$$E(L) = \mu_L \approx A_0\,\mu_C\,\mu_S^m \quad \text{and} \quad \sigma_L^2 \approx A_0^2\,\mu_S^{2m}\left[\sigma_C^2 + \left(\frac{m\,\mu_C}{\mu_S}\right)^2 \sigma_S^2\right], \qquad (4a)$$

for the expected value and the variance, respectively, in which:

$$A_0 = \left[\pi a_0 \left[F\left(\frac{\pi a_0}{w}\right)\right]^2\right]^{\frac{m}{2}} f\,\Delta t. \qquad (4b)$$

The total time interval under consideration is envisioned as a sequence of time supersteps, each subdivided into n time steps of duration Δt.

Invoking the one-dimensional theory of Brownian motion[6-8], it can be shown that the accumulated increment of crack size after a time interval $n\cdot\Delta t$ (n being an integer) has elapsed from the beginning of the j-th time superstep, x_n, is normally distributed with expected value and variance given by :

$$E(x_n) = n\cdot\mu_L \quad and \quad \sigma^2_{x_n} = n\cdot\sigma^2_L, \quad respectively \ . \tag{5}$$

Therefore, the expected value and the variance of the crack size at the end of the j-th superstep may be computed incrementally from:

$$E(a_j) = E(a_{j0}) + n\cdot\mu_L \quad and \quad \sigma^2_{a_j} = \sigma^2_{a_{j0}} + n\cdot\sigma^2_L + 2\,\gamma_{a_{j0},x_n}, \tag{6a}$$

respectively, where a_{j0} is the crack size at the beginning of the superstep and $\gamma_{aj0,xn}$ is the covariance between the random variables a_{j0} and x_n, which may be expressed as:

$$\gamma_{a_{j0},x_n} = E(a_{j0}\cdot x_n) - n\cdot\mu_L\cdot E(a_{j0}) . \tag{6b}$$

If the random variables a_{j0}, C, and S are assumed as statistically independent, the correlation term in eq.(4b) may be obtained from[7]:

$$E(a_{j0}\cdot x_n) = \pi^{\frac{m}{2}}\cdot\Delta N\cdot\mu_C\cdot E(S^m)\cdot\sqrt{\frac{\mu^{(2+m)}}{2\,\pi}}\cdot I(\mu,\nu), \tag{6c}$$

in which:

$$\mu = E(a_{j0}), \qquad \nu = \frac{\sigma_{a_{j0}}}{\mu}, \tag{6d}$$

and:

$$I(\mu,\nu) = \begin{cases} \sqrt{2\,\pi}\cdot[F(\dfrac{\pi\,a_0}{w})]^m, & j=1 \\ \displaystyle\int_{-\infty}^{+\infty}(1+\nu\,\zeta)^{(1+\frac{m}{2})}\cdot[F(\frac{\pi\,\mu}{w}(1+\nu\,\zeta))]^m\cdot e^{-\frac{\zeta^2}{2}}\,d\zeta, & j>1 \end{cases} \tag{6e}$$

Figure 1 shows a typical example of the evolution of a statistical parameter (variance) for the random

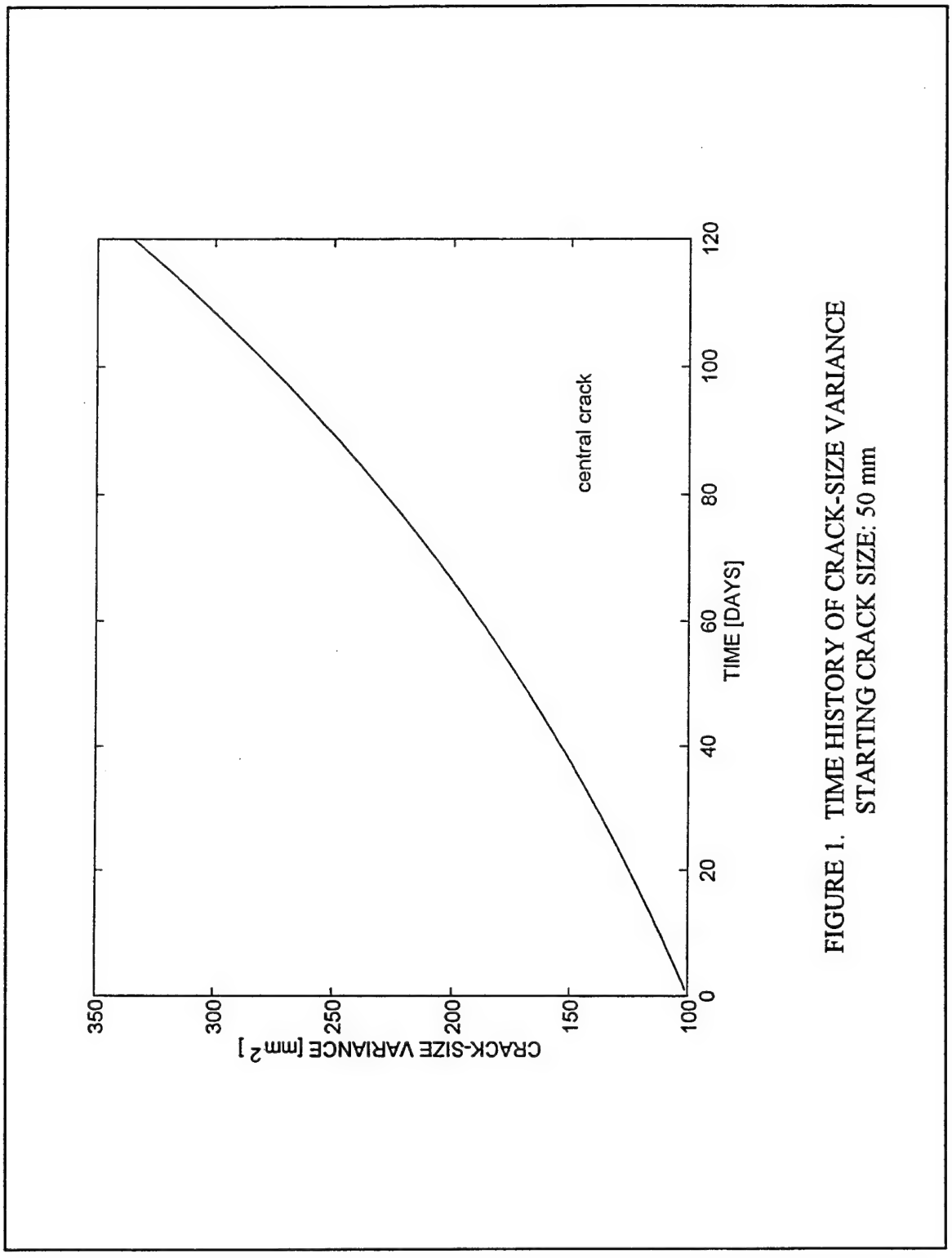


FIGURE 1. TIME HISTORY OF CRACK-SIZE VARIANCE
STARTING CRACK SIZE: 50 mm

variable crack-size (a) in time[7]. All random variables entering the formulation for this example have been assumed log-normally distributed. Notice in this example the rapid increase of the uncertainty in predictions based on the variance of the random variable. Experience with numerous examples[7] indicates that edge cracks in components of structural steel systems tend to grow at a faster rate than central cracks and, therefore, should be more critically scrutinized.

NUMERICAL CONSTRUCTION OF CRACK-STABILITY PROBABILITY CURVES

Once the estimation of the statistical characterization of the crack size is established, the analyst may automatically generate curves for the probability of not exceeding a pre-defined critical value of crack size for the metal component[9-13]. Figure 2 shows a typical example of a numerically computed curve for the probability of having a stable crack in the future. Notice the rapid reliability decay with time. Thus, reliability predictions should be updated frequently with current measures of the crack size (a_0) once the crack size actually approaches its critical value.

PROBABILISTIC DESCRIPTION OF CRACK PROLIFERATION

After the crack size reaches its critical value (inherently random), the crack proliferation (multiple splitting) is described in the light of the theory of Markovian branching processes[14]. Figure 3 shows a schematic representation of the possible crack states. Only stable crack states are considered in this investigation. It is assumed that, if the crack size exceeds its critical value, it has the potential to multiply itself until it becomes i number of cracks at a future time t, when the observation takes place (see Fig. 4); otherwise, the crack will merely increase in size during the time interval t.

At time t, the probability of still having a single crack is given by the model described above for the single-crack propagation, i.e.:

$$R_t = P(a < a_c) \tag{7}$$

On the other hand, if failure is previously defined as the event of the crack becoming at least n cracks, the probability of observing failure at time t is given by:

$$P_f = 1 - R_t - \sum_{i<n} P_i \,, \tag{8}$$

where P_i is the probability of the intersection of events $[(a>a_c) \cap$ (exactly i cracks develop at time t)].

Letting A_i be the event that exactly i cracks exist at time t , on the condition that proliferation has occurred ($a>a_c$), one has

$$P_i = P(A_i)\cdot(1 - R_t) \tag{9a}$$

And letting B_k be the event that the state ($a>a_c$) occurs at time t_k, one notes that the set of events $\{ B_k \}$ is complete and disjoint. Therefore, by the Total Probability Theorem, one has

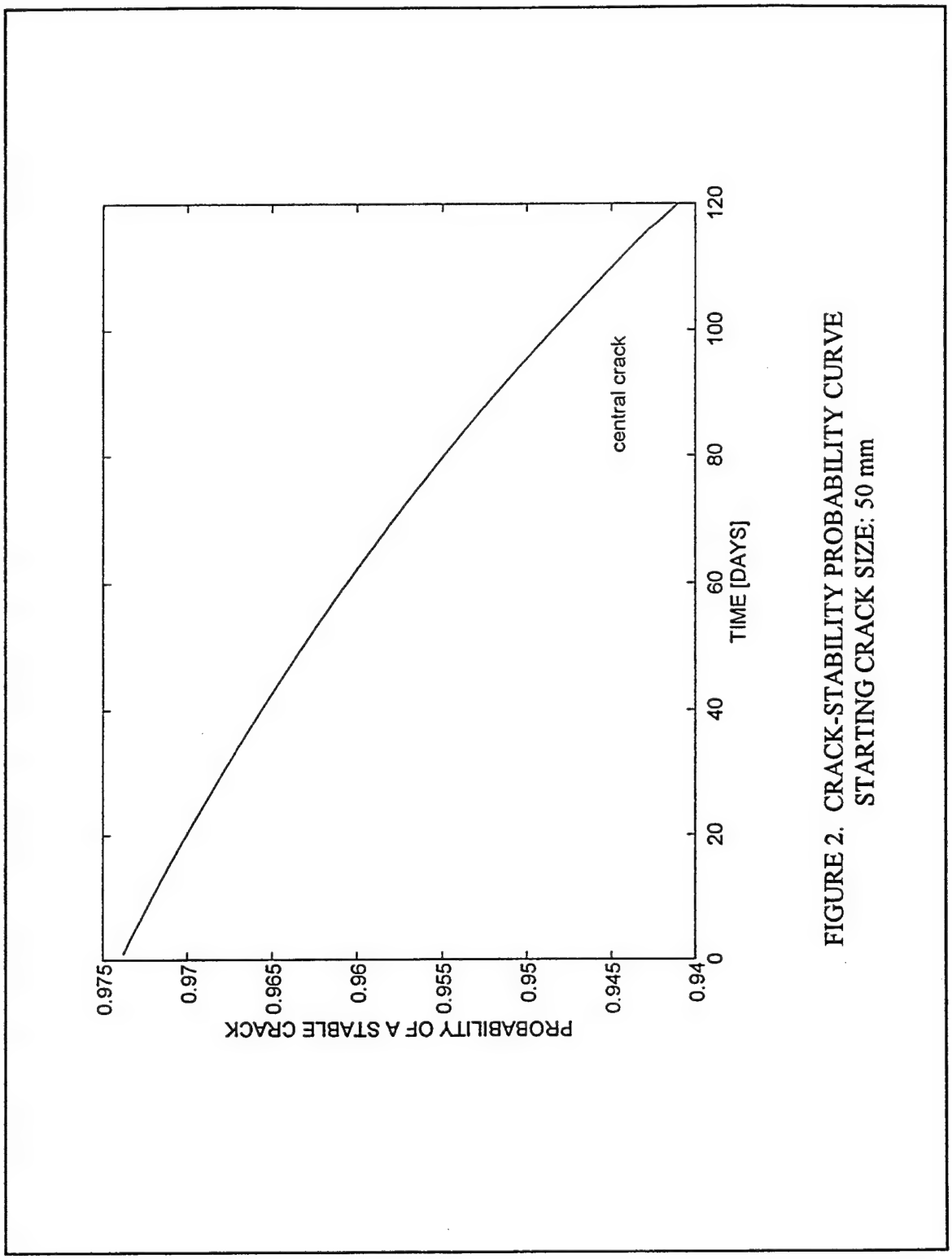


FIGURE 2. CRACK-STABILITY PROBABILITY CURVE
STARTING CRACK SIZE: 50 mm

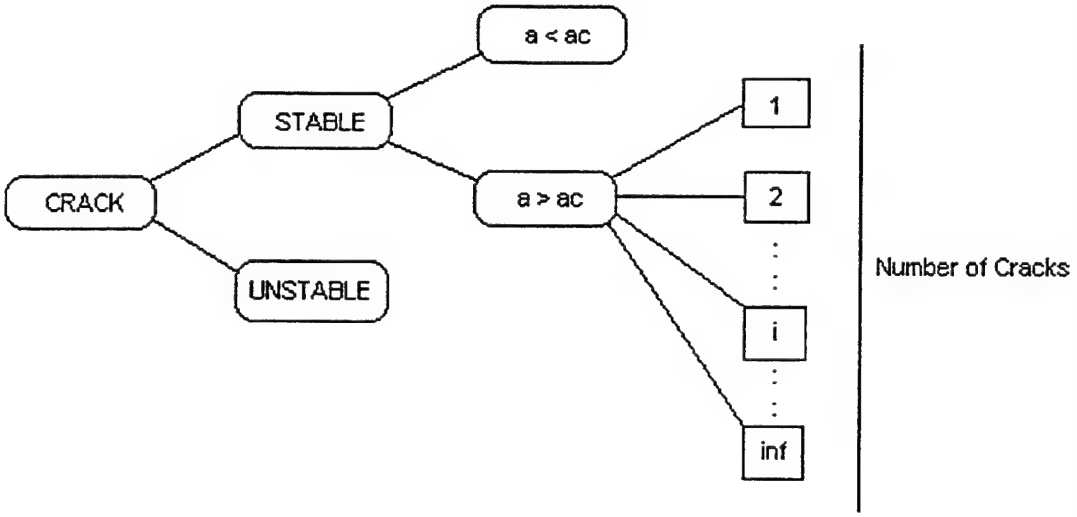


FIGURE 3. SCHEMATIC REPRESENTATION OF POSSIBLE CRACK STATES

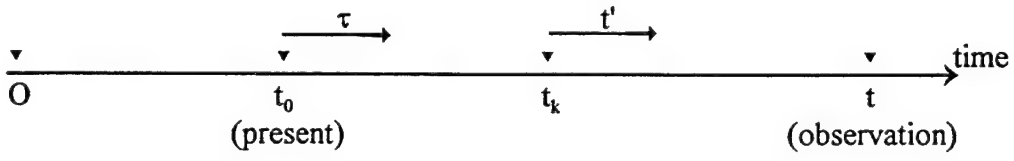


FIGURE 4. SCHEMATIC REPRESENTATION OF EVENTS IN TIME

$$P(A_i) = \sum_k P(A_i/B_k)\cdot P(B_k) \qquad (9b)$$

Invoking the Markovian assumption for the random process a(t), one may write:

$$P(B_k) = R_{k-1}\cdot(1 - R_k), \qquad (9c)$$

and, from the theory of branching processes[14], one has the conditional probability:

$$P(A_i/B_k) = p_i(t') = p_i(t - t_k) = e^{-\lambda\cdot(t-t_k)}\cdot[1 - e^{-\lambda\cdot(t-t_k)}]^{(i-1)}, \qquad i \geq 1, \qquad (9d)$$

where λ is the density of the transition out of the single-crack state into bifurcation, and crack annihilation is not possible (λ must be determined experimentally).

Substituting eqs. (9a-d) into eq. (8), one obtains an expression for the probability of failure at time t:

$$P_f = (1 - R_i)\cdot[1 - \sum_{i<n} \sum_k R_{k-1}\cdot(1 - R_k)\cdot e^{-\lambda\cdot(t-t_k)}\cdot[1 - e^{-\lambda\cdot(t-t_k)}]^{(i-1)}, \qquad (10)$$

which must be computed numerically.

CONCLUSIONS

A mathematical model is outlined to evaluate the reliability of cracked metal components under cyclic tensile stress during a specified time interval. Both centrally located and edge cracks are included in the formulation. After reaching a random critical crack size, the potential crack proliferation is described in the light of the theory of Markovian branching processes.

ACKNOWLEDGMENTS

The author gratefully acknowledges permission from the Chief of Engineers to publish this paper.

REFERENCES

1. Paris, P., and Erdogan, F. "A Critical Analysis of Crack Propagation Laws." Journal of Basic Engineering. Transactions ASME, Series D, vol. 85 (4), p. 528-534, 1963.

2. Ichikawa, M. "Probabilistic Fracture Mechanics Investigation of Fatigue Crack Growth Rate," in Statistical Research on Fatigue and Fracture. Vol. 2, ed. Tanaka, T., Elsevier Applied Science, New York, N.Y., 1987.

3. Dai, S-H., and Wang, M-O. Reliability Analysis in Engineering Applications. Van Nostrand Reinhold, NewYork, N.Y., 1992.

4. Tada, H. The Stress Analysis of Cracks, Handbook. 2nd ed., Paris Productions, St. Louis, MO., 1985.

5. Benjamin, J. R., and Cornell, C. A. Probability, Statistics, and Decision for Civil Engineers.

McGraw-Hill, New York, N.Y., 1970.

6. Clough, R. W., and Penzien, J. Dynamics of Structures. McGraw-Hill, New York, N.Y., 1975.

7. de Bejar, L.A. "Risk-Consistent Capital Allocation in Repairing Cracks in Metal Components of Hydraulic Structures." Proceedings of the 1995 Corps of Engineers Structural Engineers Conference, San Antonio, Texas, Aug. 1995.

8. Kreyszig, E. Advanced Engineering Mathematics. 6th ed., John Wiley & Sons, New York, N.Y., 1988.

9. Ditlevsen, O. "Generalized Second Moment Reliability Index." Journal of Structural Mechanics, vol. 7 (4), p. 435-451, 1979.

10. Ditlevsen, O. "Principle of Normal Tail Approximation." Journal of the Engineering Mechanics Division, ASCE 107 (EM6), p. 1191-1208, 1981.

11. Ellingwood, B., Galambos, T. V., MacGregor, T. G., and Cornell, C. A. "Development of a Probability-Based Load Criterion for American National Standard A58, Building Code Requirements for Minimum Design Loads in Buildings and Other Structures." NBS-SP-577, National Bureau of Standards, Washington, D.C., 1980.

12. Hohenbichler, M., and Rackwitz, R. "First-Order Concepts in System Reliability." Structural Safety, vol. 1, p. 177-188, 1983.

13. Shinozuka, M. "Basic Analysis of Structural Safety." Journal of the Structural Engineering Division. ASCE 109(3), p. 721-740, 1983.

14. Rozanov, Y.A. Probability Theory: A Concise Course. Dover Publications, New York, N.Y., 1977.

ATTENDANCE LIST

U.S. ARMY CONFERENCE ON APPLIED STATISTICS
16–20 OCTOBER 1995

Professor Russell Barton
Dept. of Ind. & Mfg. Eng.
207 Hammond Bldg.
Penn. State University
University Park, PA 16802

Dr. Barnard H. Bissinger
281 W. Main St.
Middletown, PA 17057

Ann E. M. Brodeen
U.S. Army Research Laboratory
ATTN: AMSRL-IS-TP
Aberdeen Proving Ground, MD 21005-5067

Dr. Marshall Brunden
Biomathematics - The Upjohn Co.
24170 Oak Lane
Mattawan, MI 49071

Dr. Aivars K. Celmins
U.S. Army Research Laboratory
ATTN: AMSRL-CI-CA
Aberdeen Proving Ground, MD 21005-5067

Professor W. J. Conover
College of Business Admin.
Texas Tech. University
Lubbock, TX 79409

Dr. Luis A. de Bejar
U.S. Army Engineer WES
Structures Laboratory, SMD-A
3909 Halls Ferry Rd.
Vicksburg, MS 39180-6199

Lou Delattre
AMSAA
ATTN: AMSXSY-RAD
Aberdeen Proving Ground, MD 21005-5071

Professor Asit P. Basu
Department of Statistics
Univ. of Missouri-Columbia
222 Math Sciences Bldg.
Columbia, MO 65211

Dr. Barry A. Bodt
U.S. Army Research Laboratory
ATTN: AMSRL-SC-S
Aberdeen Proving Ground, MD 21005-5067

Dr. Mel Brown
U.S. Army Research Office
ATTN: AMXRO-RT (70)
P. O. Box 12211
Research Triangle Park, NC 27709-2211

Robert J. Burge
Division of Biometrics
Bldg. 83
WRAIR
Washington, DC 20307-5100

Dr. Albert Chang
U.S. Army Research Laboratory
ATTN: AMSRl-MA-PD
Aberdeen Proving Ground, MD 21005-5067

Dr. David F. Cruess
Uniformed Serv. University of Health Sciences
4301 Jones Bridge Rd.
Bethesda, MD 20814

Dr. Paul J. Deason
U.S. Army TRAC-WSMR
ATTN: ATRC-WMA
White Sands Missile Range, NM 88002-5505

Mary K. Edwards
Director, AMSAA
ATTN: AMXSY-RI
Aberdeen Proving Ground, MD 21005-5071

Eugene Dutoit
Director
Dism. Battlespace Battle Lab
ATTN: ATSH-WCS
Fort Benning, GA 31905-5000

MAJ Donald W. Engen
Dept. of Mathematical Sciences
U.S. Military Academy
West Point, NY 10996-1786

Douglas H. Frank
Indiana Univ. of Pennsylvania
Dept. of Mathematics
208 Straight Hall
Indiana, PA 15705-1072

David R. Gonzalez-Barreto
Dept. of Ind. & Mfg. Eng.
207 Hammond Bldg.
Penn. State University
University Park, PA 16802

Dr. Jock O. Grynovicki
U.S. Army Research Laboratory
ATTN: SLCHE-HU/EDP
Aberdeen Proving Ground, MD 21005-5425

Andrew Harrell
Commander
USA EWES-GM
3909 Halls Ferry Rd.
Vicksburg, MS 39180-6199

Lorrie L. Hoffman
Dept. of Statistics
University of Central Florida
Orlando, FL 32186-2370

Barbara J. Kasschenbach
USACSTA
Engr. Dir.
ATTN: STECS-EN-BA
Aberdeen Proving Ground, MD 21005-5059

Paul B. Kluge
Walter Reed Army Inst. of Reg.
6445 Sewells Orchard Dr.
Columbia, MD 21045

Marc N. Elliott
1452 19th St.
Apartment 3B
Santa Monica, CA 90404

LTC John Ferguson
U.S. Army TRAC-WSMR
ATTN: ATRC-WMA
White Sands Missile Range, NM 88002-5505

Professor Jim Gentle
CSI
George Mason University
Fairfax, VA 22030-4444

Dr. John W. Green
Haskell Lab of Toxicology
P. O. Box 50
Elkton Rd.
Newark, DE 19714-0050

Linda Hall
Commander
S. S. Army Aberdeen Test Center
ATTN: STEAC-EN-BA, Bldg. 363
Aberdeen Proving Ground, MD 21005-5059

Dr. Bernard Harris
Dept. of Statistics
Univ. of Wisconsin-Madison
1210 W. Dayton St.
Madison, WI 53706-1685

Todd Jones
USA OPTEC
4501 Ford Ave, Suite 920
Alexandria, VA 22302-1458

MAJ Thomas Kastner
903 Walton Way
Smyrna, GA 30082

Kragg P. Kysor
U.S. Army Research Laboratory
ATTN: AMSRL-HR-S
Aberdeen Proving Ground, MD 21005

Dr. Robert Launer
U.S. Army Research Office
Mathematical & Com. Sci. Div.
P. O. Box 12211
Research Triangle Park, NC 27709-2211

Dr. John W. Lyons
Director
U.S. Army Research Laboratory
ATTN: AMSRL-D
2800 Powder Mill Rd.
Adelphi, MD 20783-1197

Bryson McCool
U.S. Army TRAC-WSMR
ATTN: ATRC-WMA
White Sands Missile Range, NM 88002-5505

Linda L. C. Moss
U.S. Army Research Laboratory
ATTN: AMSRL-SL-BV
Aberdeen Proving Ground, MD 21005-5066

Professor Marianna Pensky
Dept. of Mathematics
University of Central Florida
Orlando, FL 32816-2370

Professor Andrew Rukhin
Dept. of Mathematics
UMBC
Baltimore, MD 21228

Dr. Francisco Samaniego
University of California-Davis
Division of Statistics
Davis, CA 95616

Dr. Les Lancaster
Probabilistic Risk Analysis Branch
Office of Nuclear Regulatory Research
M/S NLS-372
Washington, DC 20555

Professor Wei-Yin Loh
Department of Statistics
University of Wisconsin-Madison
1210 W. Dayton St.
Madison, WI 53706

Dale A. Madden
c/o Food & Drug Admin. 10-64
5600 Fisher's Lane
Rockville, MD 20857

Craig Morrissette
13025 Broadmove Rd.
Silver Spring, MD 20904

MAJ David H. Olwell
Dept. of Mathematical Sciences
U.S. Military Academy
West Point, NY 10996-1786

Roy Reynolds
Director
U.S. Army TRAC-WSMR
White Sands Missile Range, NM 88002

Dr. Carl T. Russell
U.S. Army TEC
ATTN: CSTE-TEC-T
Fort Hunter Liggett, CA 93928-5000

Professor Jayaran Sethuraman
Department of Statistics
Florida State University
Tallahassee, FL 32306

Dr. Dennis E. Smith
Desmatics Inc.
P. O. Box 618
State College, PA 16804

Raymond Spring
U.S. Army Natick RD&E Center
Adv. Sys. Concepts Dir.
ATTN: SATNC-AAM
Natick, MA 01760-5015

Dr. Douglas B. Tang
Chief: Dept. of Biostatistics & Applied Mathematics
WRAIR
Washington, DC 20307-5100

Deloris Testerman
U.S. Army TEXCOM
ATTN: CSTE-TAV
Fort Hood, TX 76544

Albert W. Van Horn
Cost and System Analysis Dir.
U.S. Army TACOM
ATTN: AMSTN-VSM
Warren, MI 48397-5000

Clairice T. Veit
RAND
1700 Main St.
P. O. Box 2138
Santa Monica, CA 90407-2138

R. John Weaver
7000 Portage Rd.
The Upjohn Co.
Kalamazoo, MI 49001

Edward J. Wegman
Center for Comp. Statistics
George Mason University
242 Science Tech Bldg.
Fairfax, VA 22030

Dr. Madeline B. Swann
U.S. Army Research Laboratory
ATTN: AMSRL-HR-S
Aberdeen Proving Ground, MD 21005

Dr. Malcolm S. Taylor
U.S. Army Research Laboratory
ATTN: AMSRL-SC-S
Aberdeen Proving Ground, MD 21005-5067

Professor James R. Thompson
Rice University
Dept. of Statistics
P. O. Box 1892
Houston, TX 77251-1892

Dr. Mark G. Vangel
NIST
Statistical Engineering Div.
Bldg. 101/A337
Gaithersburg, MD 20899-0001

Thomas R. Walker
U.S. Army Combat Systems Test Activity
ATTN: STECS-DA-PS
Aberdeen Proving Ground, MD 21005-5059

David W. Webb
U.S. Army Research Laboratory
ATTN: AMSRL-WT-PB
Aberdeen Proving Ground, MD 21005-5066

Dr. W. Max Woods
Department of Operations Research
Naval Postgraduate School
Monterey, CA 93943

NO. OF COPIES	ORGANIZATION
2	DEFENSE TECHNICAL INFO CTR ATTN DTIC DDA 8725 JOHN J KINGMAN RD STE 0944 FT BELVOIR VA 22060-6218
1	DIRECTOR US ARMY RESEARCH LAB ATTN AMSRL OP SD TA 2800 POWDER MILL RD ADELPHI MD 20783-1145
3	DIRECTOR US ARMY RESEARCH LAB ATTN AMSRL OP SD TL 2800 POWDER MILL RD ADELPHI MD 20783-1145
1	DIRECTOR US ARMY RESEARCH LAB ATTN AMSRL OP SD TP 2800 POWDER MILL RD ADELPHI MD 20783-1145
	ABERDEEN PROVING GROUND
2	DIR USARL ATTN AMSRL OP AP L (305)

NO. OF COPIES	ORGANIZATION
1	US ARMY ENGINEER WES DR L DE BEJAR STRUCTURES LABORATORY SMD A 3909 HALLS FERRY RD VICKSBURG MS 39180-6199
1	US ARMY RESEARCH OFFICE ATTN AMXRO RT 70 DR M BROWN PO BOX 12211 RESEARCH TRIANGLE PARK NC 27709-2211
3	US ARMY TRAC WSMR DR P DEASON ATTN ATRC WMA WHITE SANDS MISSILE RANGE NM 88002-5505
3	DIRECTOR DISM BATTLESPACE BATTLE LAB ATTN ATSH WCS E DUTOIT FT BENNING GA 31905-5000
1	US MILITARY ACADEMY DEPT OF MATHEMATICAL SCIENCES MAJ D ENGEN WEST POINT NY 10996-1786
1	COMMANDER USA EWES GM ATTN A HARRELL 3909 HALLS FERRY RD VICKSBURG MS 39180-6199
1	WALTER REED ARMY INST OF RSRCH ATTN P KLUGE 6445 SEWELLS ORCHARD DR COLUMBIA MD 21045
1	US ARMY TRAC WSMR ATTN ATRC WMA LTC J FERGUSON WHITE SANDS MISSILE RANGE NM 88002-5505
1	USA OPTEC ATTN T JONES 4501 FORD AVE STE 920 ALEXANDRIA VA 22302-1458

NO. OF COPIES	ORGANIZATION
1	US ARMY RESEARCH OFFICE MATHEMATICAL & COM SCI DIV ATTN DR R LAUNER P O BOX 12211 RESEARCH TRIANGLE PARK NC 27709-2211
1	DIR USARL ATTN AMSRL D DR J LYONS 2800 POWDER MILL RD ADELPHI MD 20783-1197
1	US ARMY TRAC WSMR ATTN ATRC WMA B MCCOOL WHITE SANDS MISSILE RANGE NM 88002-5505
1	US MILITARY ACADEMY DEPT OF MATHEMATICAL SCI MAJ D OLWELL WEST POINT NY 10996-1786
1	DIRECTOR US ARMY TRAC WSMR ATTN R REYNOLDS WHITE SANDS MISSILE RANGE NM 88002-5505
3	US ARMY TEC ATTN CSTE TEC T DR C RUSSELL FT HUNTER LIGGETT CA 93928-5000
1	US ARMY NATICK RD&E CTR ADV SYS CONCEPTS DIR ATTN SATNC AAM R SPRING NATICK MA 01760-5015
3	WALTER REED ARMY INST OF RSRCH CHIEF DEPT OF BIOSTAT & APPLIED MATH ATTN DR D TANG WASHINGTON DC 20307-5100
1	US ARMY TEXCOM ATTN CSTE TAV D TESTERMAN FT HOOD TX 76544

NO. OF COPIES	ORGANIZATION
1	US ARMY TACOM COST & SYS ANAL DIR ATTN AMSTN VSM A VAN HORN WARREN MI 48397-5000
1	NAVAL POSTGRADUATE SCHOOL DEPT OF OPERATIONS RSRCH ATTN DR W WOODS MONTEREY CA 93943
1	PROBABILISTIC RISK ANAL BR OFC OF NUCLEAR REGULATORY RSRCH ATTN DR L LANCASTER M S NLS 372 WASHINGTON DC 20555
1	FOOD & DRUG ADMIN 10 64 ATTN DALE MADDEN 5600 FISHERS LN ROCKVILLE MD 20857
3	NIST STATISTICAL ENGNRNG DIV ATTN DR M VANGEL BLDG 101 A337 GAITHERSBURG MD 20899-0001
1	HASKELL LAB OF TOXICOLOGY ATTN DR J GREEN P O BOX 50 ELKTON RD NEWARK DE 19714-0050
1	PENN STATE UNIV ATTN PROF R BARTON DEPT OF IND & MFG ENG 207 HAMMOND BLDG UNIVERSITY PARK PA 16802
1	TEXAS TECH UNIVERSITY PROF W CONOVER COLLEGE OF BUSINESS ADMIN LUBBOCK TX 79409
1	UNIV OF MISSOURI-COLUMBIA DEPARTMENT OF STATISTICS PROF A BASU 222 MATH SCIENCES BLDG COLUMBIA MO 65211

NO. OF COPIES	ORGANIZATION
3	UNIFORMED SERV UNIV OF HEALTH SCI DR D CRUESS 4301 JONES BRIDGE RD BETHESDA MD 20814
1	INDIANA UNIV OF PA DEPT OF MATHEMATICS ATTN D FRANK 208 STRAIGHT HALL INDIANA PA 15705-1072
1	PENN STATE UNIV DEPT OF IND & MFG ENG ATTN D GONZALEZ-BARRETO 207 HAMMOND BLDG UNIVERSITY PARK PA 16802
1	UNIV OF CENTRAL FLORIDA DEPT OF STATISTICS ATTN L HOFFMAN ORLANDO FL 32186-2370
1	CSI GEORGE MASON UNIV ATTN PROF J GENTLE FAIRFAX VA 22030-4444
1	UNIV OF WISCONSIN MADISON DEPT OF STATISTICS ATTN DR B HARRIS 1210 W DAYTON ST MADISON WI 53706-1685
1	UNIV OF CENTRAL FLORIDA DEPT OF MATHEMATICS ATTN PROF M PENSKY ORLANDO FL 32816-2370
1	UMBC DEPT OF MATHEMATICS ATTN PROF A RUKHIN BALTIMORE MD 21228
1	UNIV OF CALIFORNIA DAVIS DIV OF STATISTICS ATTN DR F SAMANIEGO DAVIS CA 95616

NO. OF COPIES	ORGANIZATION
1	UNIV OF WISCONSIN MADISON DEPT OF STATISTICS ATTN PROF W LOH 1210 W DAYTON ST MADISON WI 53706
1	FLORIDA STATE UNIV DEPT OF STATISTICS ATTN PROF J SETHURAMAN TALLAHASSEE FL 32306
1	GEORGE MASON UNIV CTR FOR COMP STATISTICS ATTN E WEGMAN 242 SCIENCE TECH BLDG FAIRFAX VA 22030
1	RICE UNIVERSITY DEPT OF STATISTICS ATTN PROF J THOMPSON P O BOX 1892 HOUSTON TX 77251-1892
1	BIOMATHEMATICS THE UPJOHN CO DR M BRUNDEN 24170 OAK LN MATTAWAN MI 49071
3	DIVISION OF BIOMETRICS R BURGE BLDG 83 WRAIR WASHINGTON DC 20307-5100
1	DESMATICS INC ATTN DR D SMITH P O BOX 618 STATE COLLEGE PA 16804
1	RAND ATTN C VEIT 1700 MAIN ST P O BOX 2138 SANTA MONICA CA 90407-2138
1	THE UPJOHN COMPANY ATTN R WEAVER 7000 PORTAGE RD KALAMAZOO MI 49001

NO. OF COPIES	ORGANIZATION
1	DR B BISSINGER 281 W MAIN ST MIDDLETOWN PA 17057
1	MARC ELLIOTT 1452 19TH ST APARTMENT 3B SANTA MONICA CA 90404
1	MAJ T KASTNER 903 WALTON WAY SMYRNA GA 30082
1	CRAIG MORRISSETTE 13025 BROADMOVE RD SILVER SPRING MD 20904
	ABERDEEN PROVING GROUND
23	DIR, USARL ATTN: AMSRL-CI-CA, DR. A. CELMINS AMSRL-HR-S, K. KYSOR DR. M. SWANN AMSRL-IS-TP, A. BRODEEN AMSRL-MA-PD, DR. A. CHANG AMSRL-SC-S, DR. B. BODT (10 CP) DR. M. TAYLOR (3 CP) AMSRL-SL-BV, L. MOSS AMSRL-WT-PB, D. WEBB AMSRL-HR-S, DR. J. GRYNOVICKI (3 CP)
2	DIR, USAMSAA ATTN: AMXSY-RAD, L. DELATTRE AMXSY-RI, M. EDWARDS
3	USA ATC ATTN: STECS-DA-PS, T. WALKER STECS-EN-BA, B. KASSCHENBACH L. HALL (BLDG 363)

USER EVALUATION SHEET/CHANGE OF ADDRESS

This Laboratory undertakes a continuing effort to improve the quality of the reports it publishes. Your comments/answers to the items/questions below will aid us in our efforts.

1. ARL Report Number/Author ARL-SR-43 Date of Report August 1996

2. Date Report Received _____

3. Does this report satisfy a need? (Comment on purpose, related project, or other area of interest for which the report will be used.) _____

4. Specifically, how is the report being used? (Information source, design data, procedure, source of ideas, etc.) _____

5. Has the information in this report led to any quantitative savings as far as man-hours or dollars saved, operating costs avoided, or efficiencies achieved, etc? If so, please elaborate. _____

6. General Comments. What do you think should be changed to improve future reports? (Indicate changes to organization, technical content, format, etc.) _____

CURRENT ADDRESS

_____
Organization

_____
Name

_____
Street or P.O. Box No.

_____
City, State, Zip Code

7. If indicating a Change of Address or Address Correction, please provide the Current or Correct address above and the Old or Incorrect address below.

OLD ADDRESS

_____
Organization

_____
Name

_____
Street or P.O. Box No.

_____
City, State, Zip Code

(Remove this sheet, fold as indicated, tape closed, and mail.)
(DO NOT STAPLE)

DEPARTMENT OF THE ARMY

OFFICIAL BUSINESS

NO POSTAGE
NECESSARY
IF MAILED
IN THE
UNITED STATES

BUSINESS REPLY MAIL
FIRST CLASS PERMIT NO 0001,APG,MD

POSTAGE WILL BE PAID BY ADDRESSEE

DIRECTOR
U.S. ARMY RESEARCH LABORATORY
ATTN: AMSRL-SC-S
ABERDEEN PROVING GROUND, MD 21005-5067